
THE UNIVERSITY *of* EDINBURGH

# Computer Simulations of Biofilm Spatial Structure

Ellen Young


Doctor of Philosophy
The University of Edinburgh
July 2020

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

In the natural environment, bacteria predominantly grow as dense, spatially structured films, chains, mats and colonies, yet studies of bacteria until the 1990s mostly considered them as idealised individuals in well-mixed samples in suspension [1]. The subsequent shift towards studying the collective behaviour of bacteria is sometimes called a 'new paradigm in prokaryotic biology' and since then fields from clinical microbiology to biological physics have studied a huge variety of phenomenon exhibited by spatially structured bacterial populations [2–4]. The spatially structured bacterial communities I consider in this work fall into two categories: those growing on a solid substrate which provides nutrients from below, such as an agar plate, and those growing on a solid surface with nutrients diffusing down from above, such as in an experimental flow cell, which are known as colonies and biofilms respectively [1]. The importance of these surface-attached communities in the natural environment is such that it is even believed that planktonic or free-swimming microbes exist only as a means of relocating from one surface to another [1].

Depending on the context, collective structures of bacteria can be harmful or beneficial to humans humans [5]. For example, in the water industry, biofilms are used to clean untreated water via a sand filter. As raw water trickles through sand grains, bacteria feed on organic material in the water and attach to the sandgrains, clearing the water of bacteria and unwanted organic material [6]. However, biofilms can also represent a problem for the water industry when they form on the interior of water distribution systems, as they can harbour human pathogens as well as cause blockages [7]. In a clinical setting, biofilms and

colonies are known to be the primary agent responsible for chronic wounds [8] and device-associated infections, such as catheter-associated urinary tract infections and septic loosening in orthopaedic joint implants [9]. More recently, it has been suggested that biofilms form the key site of *de novo* evolution of antibiotic resistance [10]. In all cases, the central aim of research on these topics is to better understand bacterial communities in order to better control them.

Bacteria living in colonies and biofilms are capable of many collective behaviours which facilitates their role in the systems described above. Some of these behaviours are due to a phenotypic switch which occurs when bacteria grow in collective structures, and others are facilitated by aspects of collective living [2]. The phenotypic switch can cause, for example, the production of a matrix of extracellular polymeric substances (EPS) which reinforces the biofilm structure or the suppression of flagellum synthesis which might destabilise the biofilm [11]. Living collectively also enables more efficient horizontal transfer of beneficial genes [2] and facilitates communication via quorum sensing, allowing them to coordinate, for example, their response to stresses like antibiotics [12] or oxidative stress [13]. The physical interactions between the bacteria in the community, for example the friction between cells as they grow, is also known to influence the collective behaviour of bacteria [14–16]. Bacterial colonies and biofilms also show complex spatial structure, for example the complex wrinkles on *Bacillus subtilus* colonies [17] or the chiral patterns of self- lubricating bactiera [18], which both contributes to, and is affected by, the collective behaviours described above.

The aspect of collective bacterial behaviour which I choose to focus on in my thesis is the spatial structure of bacterial communities. There is a significant literature which has attempted to produce phase diagrams of biofilm and colony spatial structure, which I will outline in more detail in the next chapter [19–23]. As we shall see, many of these studies characterise biofilm spatial structure using the roughness of the biofilm interface . Surface roughness is known to control diverse characteristics such the extent of pathogen adhesion, genetic mixing and hence potential for cooperation , antibiotic penetration and the chances of fixation of antibiotic resistant mutants [24–26]. These characteristics, in turn, can feed back on the roughness of the biofilm. This feedback loop means means that phase diagrams of biofilm morphology can not only help predict the biofilm spatial structure, but other aspects of biofilm behaviour too.

The particular case of the feedback loop between biofilm structure and behaviour that I am going to focus on in my thesis is the link between macroscopic

spatial structure and genetic spatial structure [24]. By genetic spatial structure, I mean whether cells different genetic lineages are mixed in together or are spatially segregated from one another [24, 27]. The key reason for studying the genetic spatial structure is to understand its impact on evolution, because in asexually reproducing organisms which are spatially constrained by the biofilm, the diversity of the genetic material which is available to respond to selective pressures is strongly influenced by the particular mix of species and strains which are present at that particular point in time and space [24]. The genetic spatial structure is related to the macroscopic spatial structure, for example rough biofilms can separate different genetic lineages into different biofilm fingers, but in general the genetic spatial structure is not straightforwardly related to the macroscopic spatial structure and there is significant scope for further work on this topic [28].

This particular research question has been of particular interest to physicists for a number of reasons. Firstly, it is clear physical mechanisms play a crucial role in spatial structure development in colonies and biofilms [15, 29, 30]. Physical scientists have found it productive to consider the basic physics of diffusion, nutrient consumption and mechanical interactions [14], as well as the elasticity of EPS [31] and the effects of environmental conditions such as flow [32] when considering the formation of distinct biofilm or colony morphologies. In this work, I primarily focus on the physical mechanisms which influence biofilm growth, such as diffusion, mechanical interactions and cell size, rather than the models that try to include biological mechanisms such as gene regulatory changes. Secondly, the phase behaviour of biofilm morphology is also of interest to physicists because it has implications for understanding the phase behaviour of non-equilibrium systems, including non-equilibrium interface growth [21, 33, 34].

The aim of my work in this thesis is to produce a phase diagram of biofilm macroscopic spatial structure and to investigate the genetic spatial structure in each of these phases. To do this, I will use individual-based biofilm simulations, which allows me to easily tune many of the biofilm growth conditions and to accurately track the genetic relationship of all the cells. I analyse the behaviour of the interface roughness as means of characterising the macroscopic spatial structure, and produce a phase diagram for the phases present in my simulations. I then investigate the genetic spatial structure by using a number of different measures to look at the genetic diversity. I focus much of my analysis on the behaviour of the active layer, or in other words the growing layer of cells at the

top of the biofilm. As I will discuss in more detail throughout this thesis, this is because the active layer is thought to play a role in both the biofilm macroscopic spatial structure and the genetic spatial structure and it therefore provides me with a means to link these two types of spatial structure [28, 35, 36]. This approach of focusing on the active layer is rarely taken in the literature, and I use it to provide unique insights into biofilm spatial structure.

My thesis is structured as follows. I will begin in Chapter 2 by reviewing in detail the literature relevant to the questions of biofilm macroscopic spatial structure and the genetic spatial structure, so that I can situate my work in this context. Then, in Chapters 3 and 4 I will present the simulation methods I use in my thesis, beginning by outlining the open source biofilm simulation software iDynoMiCS I use in my thesis and then presenting the algorithm I developed to be used in combination with iDynoMiCS simulation software to reach long simulation times. In Chapters 5-7, I present my results regarding biofilm macroscopic spatial structure. Chapter 5 I analyse the two key quantities of my thesis, the interface roughness and the thickness of the growing layer (i.e. the active layer thickness). Chapter 6 outlines my observation of interface pinning behaviour, in which parts of the interface become stationary and lag behind the growing front, and uses this to understand the distinctive spatial structures of each of the three phases I observe. In Chapter 7 I construct a phase diagram for the phases I observe. In Chapter 8, I build directly on this work to examine the genetic spatial structure for each of the three phases. Finally, in Chapter 9 I summarise the key outcomes of my thesis and discuss interesting avenues for further work.

# Chapter 2

# Literature Review

## 2.1 Outline

In this chapter I outline the existing literature on the spatial organisation of bacterial populations and the literature on the arrangement of different cell lineages within cell populations. I begin by outlining the early experiments of bacterial colony morphology that showed that there are distinct regimes of biofilm spatial structure and inspired much of the subsequent work on biofilm and colony spatial structure. I go on to outline the interface growth theory perspective on biofilm spatial structure from statistical physics, which uses the behaviour of the roughness of the growing front to classify different regimes of growth. I also outline the reaction diffusion modelling perspective, which uses individual-based computational models and partial differential equations to elucidate the different regimes of global biofilm spatial structure. I finally outline the literature that considers the genetic spatial structure, or in other words the arrangement of different strains or cell types within the biofilm, and discuss the small literature which has attempted to link the global spatial structure of biofilms to the genetic spatial structure of biofilms.

## 2.2 Spatial organisation of bacterial communities

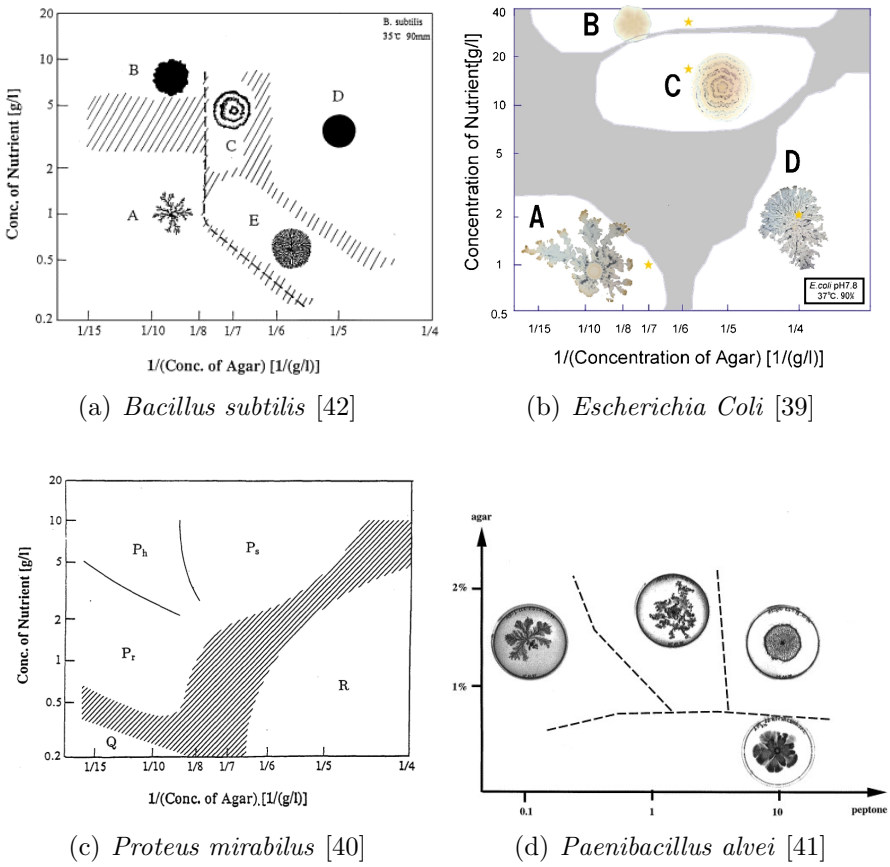
### 2.2.1 An experimental perspective

Early experiments looking at the spatial structure of bacterial colonies observed a huge range of morphologies for different experimental conditions. These experiments provided inspiration for much of the literature which is interested in phase diagrams of bacterial biofilms and colonies. The seminal experimental study in this respect is that of Matsuishi *et. al.*, who produce a phase diagram showing a variety of different *Bacillus subtilis* colony shapes for varying nutrient concentration and agar concentration (which affects bacterial motility) [37, 38]. Since then, many other experimental studies of this type have been performed, for example for *Escherichia coli* [39], *Proteus mirabilus* [40] and *Paenibacillus alvei* [41], shown in Figure 2.1. The shape of the resulting colonies vary widely from a circular compact colony for high nutrient and agar concentrations, to a concentric ring like pattern for intermediate nutrient and agar concentrations as well as colony patterns with branches of varying frequency and shape.

While most experiments in this field are on bacterial colonies growing on agar, experiments of biofilms in a flow cell also show morphological diversity. Figure 2.2 shows the set up of a flow cell. Here we can see that mature biofilms can be grown attached to the base of a flow channel, while flow of a liquid medium across the top of the biofilm can be a achieved by its connection to a microfluidic pump [43]. Biofilm flow cell experiments are more complex to perform than growing colonies on agar plates, and so most experiments in this field are performed on colonies due to their experimental tractability. Nevertheless, both smooth and biofilms in flow cells are known, with rough biofilms often being referred to as 'biofilm mushrooms' as the biofilm fingers often have significant overhangs [44]. As we shall see in the rest of our review, theoretical and computational modelling approaches very often consider the biofilm in a flow cell, in part because it presents a simpler geometry to model than that of a colony.

### 2.2.2 A statistical physics perspective.

The statistical physics literature has considered the growth of a wide variety of interfaces, including bacterial colony and biofilm growth [21, 33]. The approach




(a) *Bacillus subtilis* [42] (b) *Escherichia Coli* [39]

(c) *Proteus mirabilus* [40] (d) *Paenibacillus alvei* [41]

Figure 2.1: Experimental morphology diagrams for changing nutrient concentration (y-axis (a)-(c), x-axis (d)) and agar concentration (x-axis (a)-(c), y-axis (d)). (a) Diagram for *Bacillus subtilis*, reproduced from [42]. Phases A-E are described as diffusion-limited aggregation (DLA)-like, Eden-like), concentric ring-like, disk-like and dense branching morphology respectively. (b) Diagram for *Escherichia Coli*, reproduced from [39]. Phases A-D are described as DLA-like, Eden-like, concentric ring pattern and fluid-spreading like respectively. (c) Diagram for *Proteus mirabilus*, reproduced from [40]. Phases R, Q, Pr, Ph and Ps are described as three-dimensional growth, DLA-like, concentric ring pattern, homogeneous spreading and spatiotemporal cyclic spreading respectively. (d) Diagram for *Paenibacillus alvei*, reproduced from [41]. From left to right the phases are branching patterning, nebula patterning, stellar or galaxy patterning fuzzy branching patterning.

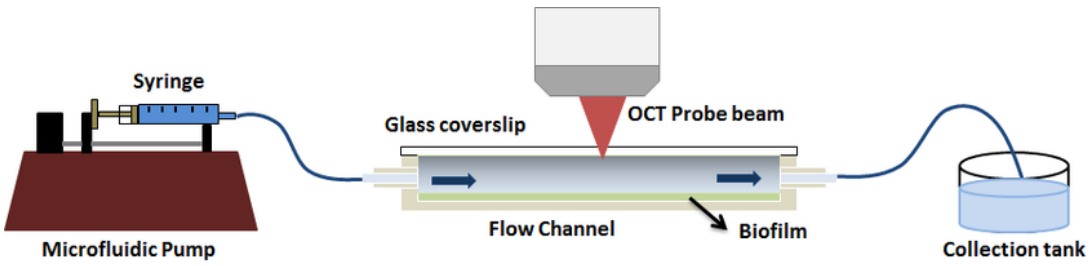



Figure 2.2: Diagram of a biofilm flow cell. Reproduced from Crusz *et. al.* [43]

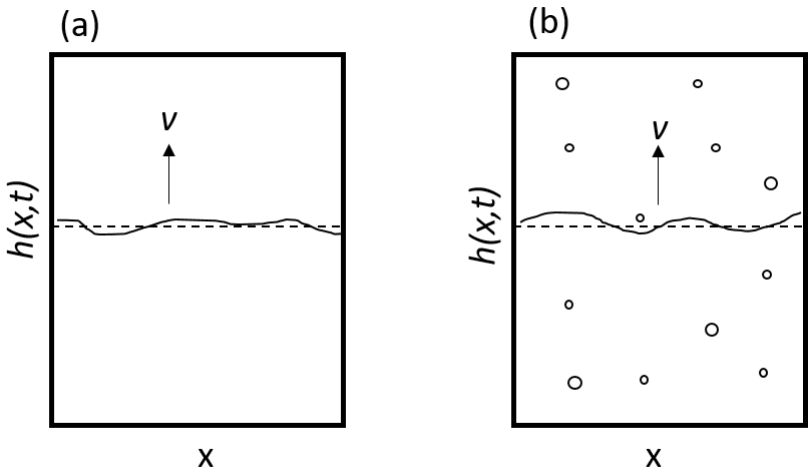



Figure 2.3: Sketches of the interface described in interface growth theory for an interface (a) with Guassian noise and (b) with quenched noise arising from inhomogeneities in the medium. The interface height $h(x,t)$ is a function of position horizontal position $x$ and time $t$. The interface is moving in a direction $v$. The circles in (b) represent the inhomogeneities in the medium. Adapted from Barabasi and Stanley [33].

of the statistical physics literature to understanding the different interface behaviours of bacterial communities is to classify their growth behaviours into universal classes based on their scaling behaviour. The aim of this is to provide a framework distinguishing interface growth patterns which may look qualitatively similar but have different physical and biological processes that produce them (though interface growth theory itself does not tell you about the actual processes). As we will outline in the rest of this section, there have been some successes in classifying the biofilm community morphology using this approach, but many questions remain unresolved.

For classifying the compact patterns (eg phases B and D In Figure 2.1(a)), the interface roughness (i.e. the surface roughness) scaling exponents are defined. We can see in Figure 2.3(a) that a one dimensional interface in interface growth theory is appropriate for a two dimensional biofilm, characterised by interface height $h$, which is a function of vertical position $x$ and time $t$ and has a linear size $L$. For this 1+1 ($x$ and $h$) dimensional biofilm, the global roughness $W$ is defined as the standard deviation of the height $h$,

$$W(L,t) = \left\langle \left[ h(x,t) - \langle h(x,t) \rangle_L \right]^2 \right\rangle^{1/2}. \tag{2.1}$$

To characterise the growth of the biofilm, it is usual to define the scaling exponents

of both the early-time and steady state behaviour. The KPZ equation obeys the Family-Viscek scaling relation [45, 46], which states:

$$W(L,t) \sim \begin{cases} t^{\beta} \text{ for } t < t_x \\ L^{\alpha} \text{ for } t > t_x. \end{cases} \quad (2.2)$$

In other words, for short times the surface roughness increases in time according to power law behaviour defined by the growth exponent $\beta$ and at time $t = t_x$ it saturates to a system size dependent value $W = W_{st}(L)$ whose finite size scaling defines the roughness exponent $\alpha$. The 'dynamical exponent' is defined as $z = \alpha/\beta$ as it determined the saturation time $t_x \sim L^{\alpha/\beta}$. The measurement of these scaling exponents is used to define the so-called 'universality class' of the interface growth.

For the compact, circular colonies, the Khadar-Parisi-Zhang (KPZ) equation was proposed from early experiments [21]. The KPZ equation is a non-linear stochastic expression for the height $h$ of an infinitely thin interface advancing at speed $v$ above a basal substratum at position $\mathbf{r}$ and time $t$ [33],

$$\frac{\partial h(\mathbf{r},t)}{\partial t} = u + D\nabla^2 h + \frac{\lambda}{2}(\nabla h)^2 + \eta(\mathbf{r},t). \quad (2.3)$$

The first term on the right represents the initial speed of the interface $u$ and the second describes the relaxation of the interface caused by the surface tension (i.e an effective diffusion term with $D$ as the diffusion constant). The third term represents lateral growth with $\lambda$ a constant which occurs because growth occurs in locally normal to the interface. The final term $\eta$ represents the system noise, which in its original form is taken to be uncorrelated Gaussian noise [33]. The scaling exponents which define this phase for the simple one dimensional interface we discussed in 1+1 ($x$ and $h$) dimensions, $\alpha = 1/2$, $\beta = 1/3$ and $z = 3/2$, though these take on other values when the system has a different geometry (eg circular colony) or number of dimensions [21, 33]. It it is now well accepted that it is not possible to measure the early time exponent $\beta$ for biofilms because biofilm formation disrupts the expected scaling behaviour [21, 47]. There has been much discussion about whether experimental results on bacterial colonies and biofilms show KPZ scaling exponents. There is some evidence that this regime may be found in a narrow range of parameters involving high nutrient concentration [35], though many authors attempting to measure the KPZ exponents have found scaling exponents larger than the expected values, meaning the interface is rougher than expected [33, 34, 48].

For compact colonies with a rough interface, the focus of the interface growth theory literature has been on the quenched KPZ equation [33, 34]. The quenched KPZ equation replaces the Guassian noise of the KPZ equation ($\eta(\mathbf{r}, t)$) with a quenched noise ($\eta(x, h(x, t))$) which is a function of $x$ and $h(x, t)$ rather than $x, t$. The quenched noise means local regions where the interface is stationary occur. The best studied source of quenched noise is inhomogenaities in the medium, which prevent the interface moving forward locally such as in Figure 2.3[34]. The quenched KPZ equation is well known to have a critical transition at a critical interface speed $u_c$, in which the quenched noise causes the interface to become completely stationary, or pinned, as a result of pinning sites hindering its forward growth. The regimes are,

- Pinned regime $u < u_c$ Pinning sites occur but they cannot be overcome and the interface growth is hindered and the interface eventually becomes stationary,
- Depinned phase $u > u_c$ Pinning sites occur but they can be overcome and the interface continues to grow,
- Flat phase $u \gg u_c$ Interface is growing sufficiently fast that all pinning sites can be overcome and the interface remains flat. This phase is equivalent to as Equation 2.3.

Authors such as Bonacehlla *et. al.* argue that the experimental evidence points to the quenched KPZ for the rough compact colonies, though there is not conclusive evidence for this and the source of the quenched noise is not clear as there are no external sources of noise such as inhomogeneities in the medium in these models [21]. In contrast to Bonacehlla *et. al.*'s finding, Gralka *et. al.* find the quenched KPZ equation useful for colonies only in the case where there are inhomogeneties added to the medium. There are also many other models of critically pinned interfaces in interface growth theory, including self-organised pinning where the pinning sites emerge from processes internal to biofilm growth, though their potential relevance for bacterial biofilms and colonies has not been investigated [49–51].

There have been a number of approaches to classifying the roughest phase of biofilm community morphology, such as Phase A in Figure 2.1(a). Some have argued that this very rough phase corresponds to the pinned phase, though the observation of a monotonically increasing or unstable interface roughness seems

to rule this out as the pinned phase mentioned above reaches a steady state as the interface becomes completely stationary [52, 53]. Many authors have argued that the roughest phase can be described by Diffusion Limited Aggregation (DLA) [47, 48, 54, 55]. Specifically, Fujikawa and Matsushita measure a fractal dimension of less than 2 ($1.716 \pm 0.008$), which is defining of this phase [33, 56]. Models which display this scaling behaviour are known to be dominated by diffusion, hence the naming of the phase as Diffusion Limited Aggregation [47]. There has been some suggestion that there may be a transition between a pinned and DLA phase that occurs as a result of overhangs forming [21, 34]. Santalla *et. al.* have argued that instead of introducing quenched disorder into the KPZ framework, introducing a shadowing model into the KPZ framework to mimic nutrient diffusion is the best model for this behaviour [52]. This shadowing behaviour is similar to the diffusion-limited aggregation models, and the shadowing effect takes its name from the simple analogy of an ensemble of grass leaves which are striving to collect sunlight: taller leaves cast shadows on shorter ones, hindering growth of the latter [52].

### 2.2.3 Reaction diffusion model perspective

Reaction-diffusion models represent an alternative, and more commonly used, approach to modelling the morphological behaviour of bacterial colonies and biofilms. Reaction-diffusion models explicitly describe the diffusion of nutrient and it's consumption during growth [57]. This is in contrast to the interface growth approach we saw in the previous section in which details of the growth mechanism are not specified explicitly. There is no universal approach to classifying phase transitions in biofilm and colony morphology in the reaction-diffusion literature. Nevertheless, as we shall see, most of these models observe at least one phase transition between smooth and rough biofilms. The processes that need to be included in these models to effectively model biofilms or colonies are the subject of significant debate in the literature [14, 16, 58]. As stated earlier, in this literature review I will discuss the physical mechanisms such as diffusion, mechanical interactions and cell size, rather than the models that try to include biological mechanisms such as gene regulatory changes.

Reaction-diffusion models of bacterial communities are most commonly either systems of partial differential equations or hybrid individual-based computational models (IBMs). In either case, they contain an explicit equation for the diffusion

of nutrients that is coupled with either continuum equations describing the bacteria or the rules governing behaviour of individual agents in a simulation. Models that represent both bacteria and nutrient as continuum fields can be split into two broad categories. The first type of model only considers the interaction of the biofilm or colony interface with the nutrient field, and the biofilm or colony is described only by a single variable, the height (or colony radius) [59–61]. Other continuum models represent the biomass of the biofilm or colony via a density field equation that is coupled to the equation for the nutrient field. The biomass density field allows for the representation of spatial heterogeneity below the interface in terms of pressure and nutrient gradients. This means the interface can have a finite thickness, representing the finite thickness of the growing layer of cells at the growing edge of biofilm or colony [60].

Individual-based models instead represent bacteria as discrete objects, while still representing the nutrient as a continuum field [57]. In these individual-based models, the collective effect of the action of each individual determines the properties at the population level including the morphology of colony or biofilm. This means that feedbacks between the behaviour of individuals and the population as a whole emerge automatically rather than being included explicitly as with the wholly continuum approach. For example, a local region of space with a high density of individuals will automatically deplete nutrients more than a low density region. In earlier versions of such models, the agents representing the microorganisms were confined to a lattice and limited in how far they could move. However, more recently lattice-based models have been somewhat discredited and the increasing computational power means models where all the agents are represented in continuum space have come to the fore [62]. There are a variety of such models have been proposed, including iDynoMiCS which I use in this thesis and describe in detail in Chapter 4.

Nutrient limitation is perhaps the best studied of all the mechanisms impacting biofilm morphology. A seminal piece of work in this respect is the 2001 study by Dockery and Clapper [58]. They propose a continuum model where the growth of a one dimensional front is proportional to the local nutrient concentration alone. They use linear stability analysis to show that this biofilm interface is unstable to perturbations caused by uneven nutrient distribution, and so argue that nutrient limitation can lead to a fingering instability in the growing biofilm front [58]. Indeed, the importance of the nutrient concentration and diffusion limitation has been recognised in most studies since the early work of Matsushita *et. al.* [63],

although many authors argue though many argue that additional mechanisms are important [44, 53]. In my work I also find a crucial role for nutrient limitation in biofilm morphology.

Many authors have noted that in a growing biofilm, nutrients does not penetrate to all the cells in the biofilm, and there is often a finite growing layer of cells at the edge of the biofilm or colony [28, 42, 64]. Several authors who use continuum models representing the bacteria as a density field equation argue that the nutrient concentration controls the thickness of the active layer, which in turn is enough to explain transitions in the morphology of biofilms and colonies [42, 64]. Active layers also emerge in individual based models [28, 65].

An important work on the active layer, that I draw on significantly in my work, is that of Nadell *et. al.*. The use an IBM to predict that it is not the nutrient concentration alone which determines spatial structure, but the combination of different factors which determine the active layer thickness. This IBM is an earlier generation of biofilm IBM than I use in my work, though it is similar in it's approach to describing nutrient diffusion and cell interactions [66]. They use an analytical technique from chemical engineering to combine the factors that influence active layer thickness in their simulations into a dimensionless number $\delta$, though the details of how this is done is not described. For their simulations, $\delta$ is given by,

$$\delta = \sqrt{\frac{S_{bulk} DY}{\mu_{max} \rho h^2}} \tag{2.4}$$

where $S_{bulk}$ is the bulk liquid concentration of growth substrate, $D$ is the growth substrate diffusion coefficient, $Y$ is the yield with which cells convert substrate to biomass, $\mu_{max}$ is the maximum specific cell growth rate, $\rho$ is the cell biomass density, and $h$ is the height of the diffusion boundary layer [28]. They argue that changing any of these parameters would have the same impact on the active layer thickness and hence the spatial structure, though this hypothesis has not been systematically tested. Instead, their focus is on demonstrating that $\delta$ predicts the segregation of different cell linages during cell growth, which I will discuss further in the next section. Importantly, their work hints that the reason $\delta$ and lineage segregation are related is because of the spatial structure and differing roughnesses of the biofilms that result from the active layer behaviour, an insight which I expand on in this work.

More recently, the role of mechanical interactions in biofilm growth has been

more closely considered, leading some to argue that mechanical interactions are crucial in determining biofilm spatial structure. The mechanical interactions arise from cells pushing each other out of the way as they grow. This approach was pioneered by Volfson *et. al.*, who use experiments and a combination of hybrid and continuum models of colony growth to investigate the orientation of rod-shaped cells in growing populations. They find a transition between rod cells being randomly orientated and being aligned in their orientations that was driven by short-range mechanical interactions between the cells [14]. Smith *et. al.* later built on this work by studying the mixing of cells with different cell shapes and sizes using IBMs and experiments, finding that cell types strongly sort by shape, with round cells at the top of the colony and rod-shaped cells dominating the basal surface and edges [67]. Other authors focus on the elasticity of cells and the pressure which builds up as a consequence of the cells pushing against each other. Farrell *et. al.* argue that these interactions are responsible for a transition between smooth and rough biofilms [16]. Ghosh *et. al.* go on to argue that the elasticity of the extracellular matrix can also contribute to a similar transition [31].

Most models for biofilm and colony spatial structure do not consider the motility of cells. This is because in the biofilm mode of growth, bacterial motility is limited; instead an extracellular matrix sticks bacteria to each other and to the surface. However, there have been some studies looking at the impact of the motility of cells on biofilm spatial structure. In the early experiments we saw in Section 2.2.1, the concentric rings have been found to arise from the motility of the cells. Perhaps most importantly, it is thought that the rough 'mushroom-shaped' biofilms seen in biofilm reactors could to motile cells 'climbing' onto the top of the existing biofilm structure [44].

## 2.3 Genetic Spatial Structure of bacterial communities

In this thesis I also investigate the arrangement of different cell types within a biofilm, or as I will refer to it in the context of this thesis, the genetic spatial structure of the biofilm. We can see an example of the different kinds of genetic spatial structures that occur in simulated biofilms in Figure 2.4. In some of these biofilms, the different strains (represented by the different colours) are well

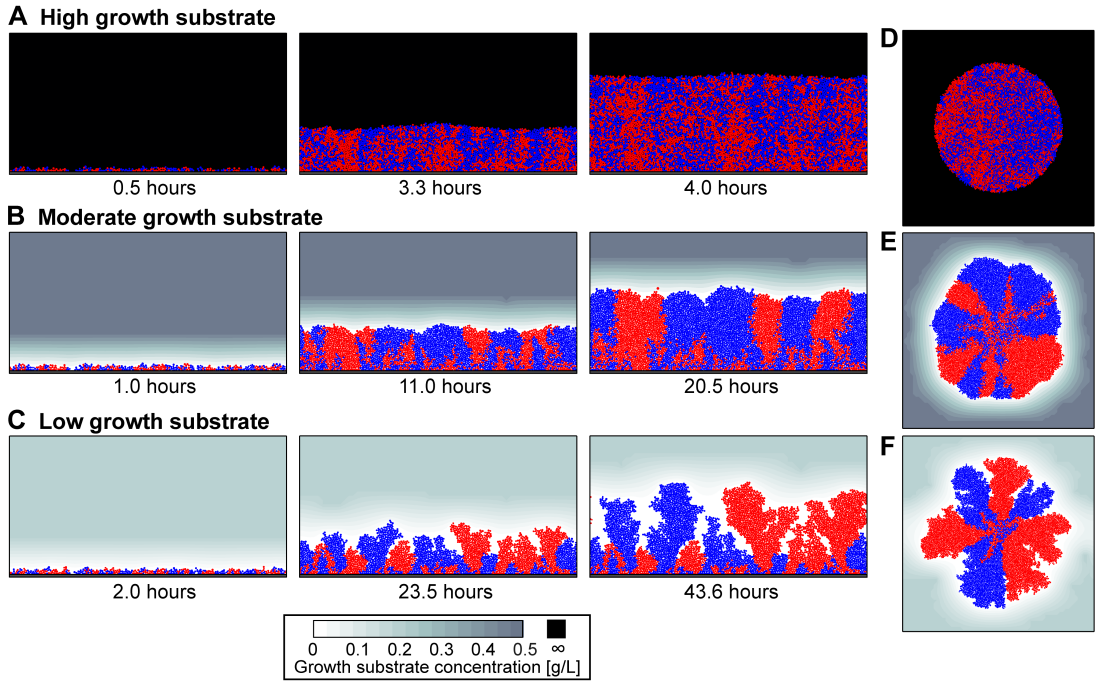



Figure 2.4: This is an example of the different genetic spatial structures which can be present in a simulated biofilm. Reproduced from Nadell *et. al.* [28].

mixed and so the biofilm can be thought of as having low genetic spatial structure. However, in some of the other biofilms, the strains are segregated from one another and so the biofilm can be thought of as having high genetic spatial structure. This distinction between low and high genetic spatial structure can be made independently of what the global spatial structure or overall morphology of the biofilm looks like. In the literature, the primary motivation for studying the genetic spatial structure has been to understand its impact on evolution, because in asexually reproducing organisms which are spatially constrained by the biofilm, the diversity of the genetic material which is available to respond to selective pressures is strongly influenced by the particular mix of species and strains which are present at that particular point in time and space [24] In what follows I will describe the particular questions about genetic biofilm spatial structure which are addressed in the biofilm literature. I will pay particular attention to attempts to understand the relationship between the genetic and global spatial structure of the biofilm, which perhaps surprisingly is not well understood.

Within this field, research has primarily focused on two specific questions. The first question concerns the degree of mixing of different strains and how it can influence evolution of cooperative and competitive traits. Hamilton's principle states that evolution can favour selection of cooperative traits when organisms in

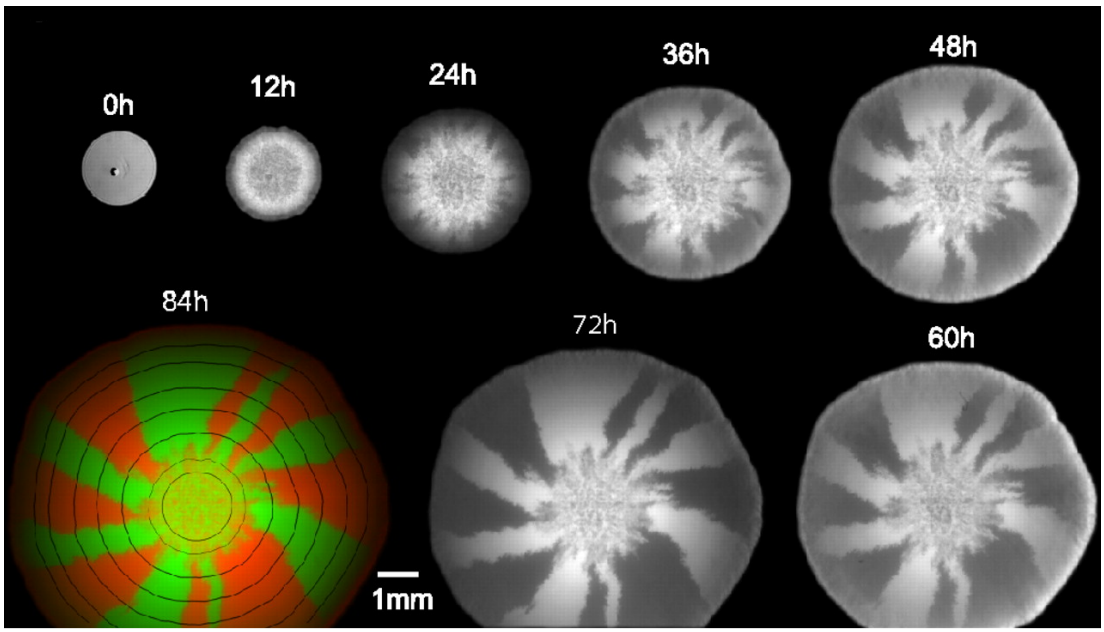



Figure 2.5: Images of bacterial colonies grown from a 50:50 mixture of red and green labeled cells showing the spatial segregation, or 'sectoring', of cells which are otherwise genetically identical. Reproduced from Hallatscek *et. al* [35].

a population are closely related to each other [68]. Therefore, this theory suggests cells in a colony or biofilm which has a highly segregated genetic spatial structure are more likely to be next to closely related cells and so more likely to develop cooperative traits than cells in a well mixed biofilm, whose neighbors would on average be less closely related [24]. An example of a cooperative trait is the production of adhesins which help the cells stick together better. An example of a competitive trait is the production toxins that damage competing cells [24, 69]. A great deal of the research on this question focuses on whether particular traits are competitive or cooperative which conditions favour the evolution of cooperative or competitive traits [26, 70, 71], though in this literature review I shall focus on the studies of the mechanisms which produce the mixing or segregation of different strains.

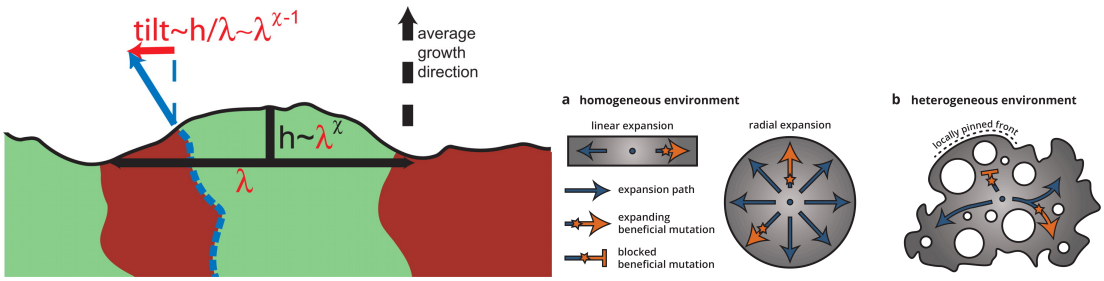
The second focus of this field has been on 'gene surfing' in expanding populations, including biofilms [35]. In the classical picture of evolution, for large, well-mixed populations, it is assumed that advantageous mutations are rare, and that each mutation 'sweeps' through the population in a deterministic manner until that mutant becomes the dominant strain [72, 73]. In the ideal case of a planar growing front (eg smooth biofilm), this would result in the fittest strain 'taking over' the whole of the growing front of the biofilm. However, subsequent research has shown that this picture does not always hold, particularly in spatially constrained

populations like biofilms and colonies [74]. For example, it is known that spatial structure can create population bottlenecks, which increase stochasticity, or 'genetic drift', such that a neutral or deleterious mutation can come to dominance [35, 36]. In practice, this phenomenon is primarily studied in colonies, in which it has been observed that several sector-like regions of the dominant strains are seen instead of a single dominant strain because of their circular geometry, such as Figure 2.5. In these studies, sectors appear even when all the strains are of equal fitness (neutral), and in my my thesis I will also focus on neutral strains.

As we have seen, the processes that influence genetic spatial structure, even for neutral strains, are relevant to both the cooperation/competition literature and the gene surfing literature. One relevant aspect of this is the role of the growing layer. Evidently, only cells which are in the growing layer can reproduce, and simulations and experiments have indicated that this reduces the effective population size, producing a spatial bottleneck [75–77]. Thus, the finite size of the active layer introduces an additional sampling effect on the population, increasing genetic drift [35, 78]. It is also thought that well-mixed populations only occur when most of the cells are in the active layer, which means all cells can continue to produce progeny and their genetic lineages are not lost [28].

The global roughness of the biofilm, which is closely related to the active layer thickness, also plays a role in genetic spatial structure. Nadell *et. al.* argues that the active layer thickness is related to the 'segregation index', or the degree to which different cell lineages are separated [28]. They argue that this is because the active layer thickness controls the interface roughness. Although they are not able to provide a quantitative relationship between the interface roughness and segregation index or the active layer thickness, their work suggests three regimes of roughness which see in Figure 2.4. For a very thick active layers (or a large nutrient concentration) there are high levels of mixing. In this regime, all cells are growing and so there is no active layer bottleneck effect. Then, there is the regime where the cell lineages are segregated but the colony remains compact. Finally, the very rough biofilms which separate the different cell lineages into different 'fingers', which then places constraints on the ability of different species to compete with different strains [28, 79].

Though there remain challenges in linking the genetic spatial structure to the global interface roughness, progress has been made in the case of compact colonies where 'sectoring' occurs. A seminal study by Hallatscek *et. al.* looks at the domain boundaries between colony sectors for two strains which are identical




(a) (b)

Figure 2.6: Subfigure (a) is reproduced from Hallatscek *et. al.* [35]. The roughness of the colonization front (black line) influences the wandering of locally perpendicular domain boundaries (red lines). In this sketch, the middle domain boundary (dashed blue) tends to follow the blue arrow indicating the local growth direction of the colony. Because of the stochastic surface growth, the local growth direction deviates from the average growth direction (black dashed arrow). Consequently, the domain boundary is subject to a drift (red arrow) transverse to the average growth direction, which is proportional to the local tilt of the interface. Subfigure (b)is reproduced from Gralka *et. al.* [80]. Sketches showing that in a homogeneous environments, a locally established beneficial mutation (orange star), can expand freely (arrows), while in heterogeneous environments, a beneficial mutation can become trapped in pinned stretches of the population.

apart from the fluorescent protein which they express [35]. It is useful to track the domain boundaries because a sector expands if its boundaries drift apart, but is annihilated if its boundaries meet. Using Figure 2.6(a), they argue that the path of these domain boundaries is determined by the position of the interface at the point where the domain boundary meets it, meaning the sectoring behaviour is determined by the local roughness of the growing front. As we mentioned in Section 2.2.2, Hallatscek *et. al.* argue their findings are consistent with it being possible to describe compact biofilms with the KPZ equation (Equation 2.3), though they do not measure the relevant scaling exponents explicitly.

Gralka *et. al.* build on the work of Hallatscek *et. al.* to study the impact of environmental heterogeneities on domain boundaries between sectors and the roughness of the expanding front [80]. They find that environmental heterogeneities caused by a patterned agar surface can locally slow down, or 'pin', the moving front, and this in turn impacts the behaviour of the domain boundaries. Where a pinning site occurs, the bacterial strains present at that point are blocked from further expansion in the population, as we see in Figure 2.6(b). This limits the number of individuals that have access to expansion paths and so increases genetic drift [80]. Gralka *et. al* observe pinning-like behaviour in their experiments, and relate this to the interface growth theory approach

outlined earlier in the chapter. They are able to measure roughness scaling exponents consistent with a universality class closely related to the quenched KPZ class outlined earlier (the quenched Edwards Wilkins class), which also has a pinning transition.

Several authors have considered the impact of the mechanical interactions between cells on the mixing of different strains. Kan *et. al* have looked at the effect of adhesive intercellular interactions at domain boundaries between different strains. They show that these adhesive interactions can determine the degree of mixing [27]. Specifically, they find that the friction caused by intercellular adhesion elongated the fractal-like boundary between cell lineages by increasing rotational motion during colony growth [27, 81]. Farrell *et. al* also show that the probability that a faster-growing mutant 'surfs' at the colony's frontier and creates a macroscopic sector depends on physical properties of the cells, such as the shape, elasticity and friction [82].

## 2.4 Summary

In this chapter, I have provided an overview of the literature on bacterial community global spatial structure and genetic spatial structure. I have specifically focused on the work which considers the physical mechanisms which impact biofilm and colony morphology, such as diffusion, mechanical interactions and cell size, rather than the models that try to include biological mechanisms such as gene regulatory changes, as this is most rel event for my work which follows. I described the early experiments on bacterial colonies on agar which demonstrated the variety of colony mythologies that existed and inspired much of the other work I consider, as well as experiments on biofilms performed in flow cells. I then outlined how the interface growth theory approach use interface roughness to classify bacterial community morphology. I finally outlined the literature that is interested in the arrangement of different strains of cells within bacterial communities, and described the approaches that have been taken to understand how these arrangements arise. In the rest of this thesis, I will outline my own work and discuss how it contributes to the literature outlined here.

# Chapter 3

# Simulation Methods A: The iDynoMiCS biofilm simulation software.

## 3.1 Outline

As we have seen in the literature reviewed in Chapter 2, individual-based modelling (IBM) is one of the key methods used in biofilm research. In these type of simulations, the bacteria which make up the system and their interactions are simulated directly, and the emergent behaviour of the group is observed. More generally individual-based, or agent-based, simulations are used across many areas of research, from modelling traffic flow to the spread of disease epidemics, following the advent of modern computing which has allowed the simulation of sufficiently large numbers of agents to be able to observe the emergent behaviour.

As with all models, IBMs involve simplifying assumptions. These assumptions ensure that the basic mechanisms which control the behaviour of the system can be elucidated, and in addition are often necessary to make the simulations computationally tractable. IBMs and other models are good for establishing the basic systems principles to be tested, though comparison with experiments is needed to establish the validity of the assumptions made and the emergent behaviour produced. Additionally, as IBMs model the behaviour of the individuals directly, these types of models often use many parameters which

are hard to calibrate unless a great deal is known about the behaviour of the individuals.

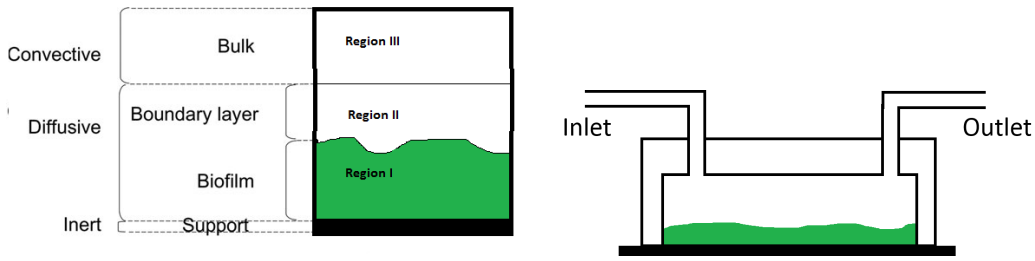
The IBM simulations in this thesis use the open-source biofilm modelling software iDynoMiCS. Briefly, iDynoMiCS models the bacteria in a biofilm as individual agents whose behaviour is coupled to a solute reaction-diffusion equation [83]. The agents, which are assumed to be discs in 2D space or spheres in 3D space, exist in continuum space, growing until they reach a maximum radius before dividing and interacting with one another via a shoving algorithm, while the solute is represented by a concentration field which varies in space and time according to diffusion and consumption by the bacteria. The simulation software does not include many of the biological interactions between cells, and assumes that the nutrient remains in a pseudo-steady state throughout the simulation.

In the rest of this chapter, I describe each of the key aspects of the iDynoMiCS simulation software in more detail: the computational domain, the agents and their interactions, the solute field and reactions and the way in which genetic information is recorded in the simulation. I will detail the assumptions made by the software and I will outline how I set up the software for use in the rest of this thesis, including the input parameters I use.

## 3.2 IBM Simulations with iDynoMiCS

### Computational Domain

The computational domain is is split into three sections, as in Figure 3.1(a). The biofilm matrix (Region I) is composed of microbial cells and is bounded by the solid support to which the agents can adhere. The bulk compartment (Region III) represents the well-mixed bulk liquid which remains at the nutrient concentration set in the input file ($S_{bulk}$). Finally, the diffusion boundary layer (Region II) is the compartment in which the nutrients diffuse down from the bulk compartment to the biofilm. A variety of boundary types are possible using iDynoMiCS, but in my simulations I choose a no-flux boundary between Region I and the support, the side boundaries are periodic so each side can interact with each other and the top boundary of Region II is set to be constant at the bulk solute concentration $S_{bulk}$. In common with most of the computational work in this field, I restrict our simulations to two dimensions for reasons of computational feasibility.




(a) Computational Domain (b) Flow Cell

Figure 3.1: Figure (a) shows the computational domain of iDynoMiCS, including the three regions mentioned in the text. Figure (b) is a sketch of a flow cell, which is modelled by our set up of iDynoMiCS.

This set up is intended to be similar to an experimental flow cell, which as we have seen in Chapter 2 is often used to study biofilms in the lab. Figure 3.1(b) shows a sketch of the flow cell set up. We can see that the biofilm grows on the base of the experimental flow cell, and nutrients flows through the flow cell from one side to the other. In the iDynoMiCS simulations, the convective flow is not modelled directly, but rather we assume there is a stationary layer of fluid close to the biofilm (the boundary layer) [62, 83]. This means we are assuming laminar flow of the nutrients in the flow cell, so we can reasonably assume that the nutrient field changes with diffusion only in a small region above the biofilm (ie the boundary layer).

### Solute Field

In the bulk compartment (Region III), the nutrient concentration is fixed at the bulk nutrient concentration $S_{bulk}$. In the diffusion boundary layer and the biofilm compartment (Regions I and II), the the solute is represented by a concentration field $S$ which varies in space $\mathbf{x}$ and time $t$, according to the equation,

$$\frac{\partial S(\mathbf{x})}{\partial t} = \nabla \cdot \left(D_S(\mathbf{x}) \cdot \nabla S(\mathbf{x})\right) - r_s(\mathbf{x}) \tag{3.1}$$

where $r_S$ is the consumption rate of the solute by the bacteria and $D_S$ is the solute diffusion coefficient. $D_S$ depends on the spatial position vector $\mathbf{x}$ as we assume the diffusivity within the biofilm is a fraction of the diffusivity outside it. Equation 3.1 any change in time of the solute field ($\partial S(\mathbf{x})/\partial t$) is determined by diffusion of the solute ($\nabla \cdot (D_S(\mathbf{x}) \cdot \nabla S(\mathbf{x}))$) and consumption by the bacteria ($r_s(\mathbf{x})$). In practice, the time scale of solute diffusion is much faster than nutrient consumption [83].

Therefore we can assume the solute field is in a pseudo-steady state compared with the agent dynamics, because the nutrient steady state is reached very rapidly in comparison to the timescale of bacterial growth (i.e. changes in the nutrient consumption term) [83–85]. This simplifies Equation 3.1 to

$$0 = \nabla \cdot \big(D_S\big(\mathbf{x}\big) \cdot \nabla S(\mathbf{x})\big) - r_s(\mathbf{x}). \tag{3.2}$$

In practice, these equations are solved using the multigrid algorithm [86–88]. This uses a hierarchy of discretisations and is often used for solving differential equations where there are multiple scales of behaviour. Briefly, the multigrid method accelerates the convergence of a basic iterative method by switching between solving the problem on fine and coarse grids. This means convergence can be achieved more rapidly than using a fine grid only. It also also means that different scales of errors, corresponding to the different scales of behaviour in the system, are accounted for. In practice, this means storing each field (e.g.$S(\mathbf{x})$) on multiple grids of different resolution [83]. Since the bacteria are modelled as individual agents, the relevant properties of the bacterial agents for the solute equations are projected onto a grid, such that each grid square contains the average of the agent's values within that grid square. For example, the rate of solute consumption $r_S(\mathbf{x})$ depends upon the growth rate of the bacteria at each position and so a grid of averaged bacterial growth rates is produced for the purposes of solving this equation.

### Bacterial Agents

As I mentioned, the bacteria are modelled as individual agents, with a number of rules governing their behaviour. The initial number of bacteria are specified at the start of the simulation, and their initial positions are randomly chosen on the base of the biofilm reactor (see Figure 3.1(a)). The biomass density of the bacterial cells $\rho$ is constant, but the initial radius is randomly chosen around an average value, in order to better model biological variability [83]. The agents then grow and divide in response to the solute field, and interact mechanically with each other over the course of their growth, the details of which I will specify below. The bacteria are not motile other than movement from growth and from mechanically interacting with one another. While some motility is seen in biological biofilms, this motility is often small (see Chapter 2) and the main features of biofilm growth can be captured using non-motile agents.

The bacterial agents grow in response to the solute availability of the local environment. The growth rate $\mu$ of each bacterium is governed by the well known Monod function,

$$\mu = \mu_{max} \frac{S}{k_S + S}, \tag{3.3}$$

where $\mu_{max}$ is the maximum specific growth rate of the bacteria, $k_S$ is the concentration of the solute at which the growth is half maximal, and $S$ is the local solute concentration of the bacterial cell. We saw earlier that the solute field is stored in a grid, and so $S$ is the solute concentration of the grid square the cell is part of. The amount of biomass which is produced with the consumption of the solute is determined by the yield constant $Y$, which is the grams of biomass produced for the grams of solute consumed.

The bacterial agents reproduce asexually. Once the agents have reach a threshold division radius, they divide into two daughter cells. The daughter cells are positioned with zero overlap and equidistant from the mother cell's centre, but are oriented in a random direction. As with the other properties of the bacteria, the division radius, radii of the daughter cells and direction of orientation are chosen randomly about mean values which are specified at the start of the simulation.

iDynoMiCS approximates mechanical interactions between the cells using a relaxation (or 'shoving') algorithm on the local level of the agents. As we can see in Figure 3.2, each cell has a defined total radius $r_i^{Tot}$ and shove radius $r_i^{Shov}$, where $r_i^{Shov} = k_{Shov} r_i^{Tot}$. The shove radius $r_i^{Shov}$ can be thought of as the cells radius of influence and $k_{Shov}$ controls the degree of packing of the agents($k_{Shov} > 1$). The spheres of influence (defined by $r_i^{Shov}$) of two cells with centers distance $d_{1,2}$ apart overlap by an amount $\delta$, such that

$$\delta_{1,2} = k_{Shov}(r_1^{Tot} + r_2^{Tot}) - d_{1,2}. \tag{3.4}$$

In each agent time step, any growth in the agents' radii is calculated. After this, each agent is visited one by one in a random order and checked for overlaps with its neighbors. For any pair of cells with any overlap $\delta_{i,j}$, each of the agents are shifted by a distance $\delta_{i,j}/2$ along the vector of their centres $d_{1,2}$, such that they no longer overlap. However, these agent re-locations are not applied immediately, but are summed and the net movement applied after all movements has been calculated. This process is repeated until a minimised overlap state is found [83].

One of the strengths of iDynoMiCS is that it can track the genealogy of cells

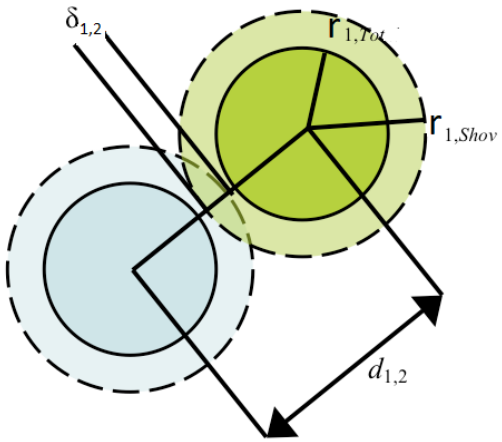



Figure 3.2: This Figure shows two agents in an iDynoMiCS at the point at which they would be 'shoved' apart. $\delta_{1,2}$ is the overlap between agents, $d_{1,2}$ is the distance between the centres of the two agents, $r_1^{Tot}$ is radius of the agent which can not be overlapped by another agent and $r_1^{Shov}$ is the shoving radius, such that a distance $r_1^{Shov} - r_1^{Tot}$ may be overlapped by other agents. Reproduced from Lardon *et. al.* [83].

within the biofilm. There are three numbers associated with each cell that together fully describe the how the cells are related to one another - the family, the generation and the genealogy. A unique family number $F_n$ is assigned as $1...n$ for the $N_0$ initial cells that are present at the start of the simulation. The progeny of the initial cells retain the same family number, so each initialised cell produces a genetic tree such as that in Figure 3.3. At the start of the simulation, each initial cell is assigned a generation $D_n$ of zero. The generation number then increases by one each time a cell divides, and both of the daughter cells inherit the same generation number (shown in red in Figure 3.3). To uniquely define each cell's position in the genetic tree, one further number is required - the genealogy $G_n$. Each initial cell is assigned a genealogy of zero. Then, when a cell divides, one cell retains the genealogy of the mother cell $G_{mother}$, while the other gets a new genealogy $G_{new}$, as,

$$G_{new} = 2^{D_{new}-1} + G_{mother} \tag{3.5}$$

where $D_{new}$ is the current generation (i.e. the generation after division). This procedure, shown in Figure 3.3 (black numbers), means that each cell's position in the family tree is uniquely specified.

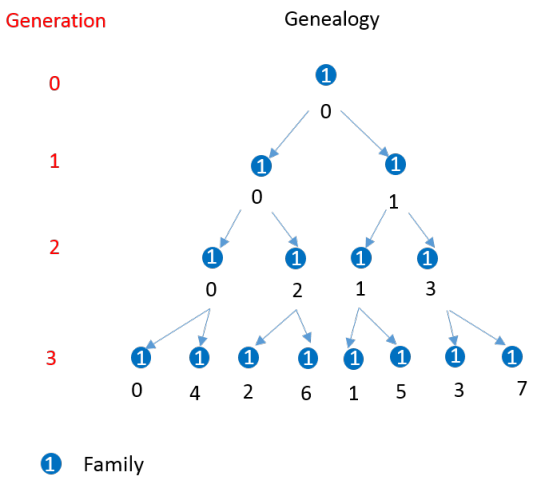



Figure 3.3: This Figure shows a sample genetic tree beginning from a single cell. The generations of the cells are shown on the left in red. The genealogy of each cell is shown in black below each cell. The family number of each cell is shown in white on top of the cell.

### iDynoMiCS Algorithm

Together, the reaction-diffusion equations and the behaviour of the bacterial agents which describe biofilm growth are implemented via Algorithm 1 [83]. We can see that in step 1 the simulation begins by initialising the positions of the starting agents, either from a specified configuration or by randomly distributing a specified number of agent along the base. In step 2, the steady-state solute concentrations are calculated from the boundary conditions and distribution of agents with their reaction rates. Then in step 3, a number of global time steps are implemented until the total simulation time has elapsed. Within each global time step, there are a number of agent time steps in which the processes the agents undergo are implemented one after another. After this, the steady-state solute concentrations are calculated for the new configuration of cells. This means the solute concentration field dynamics and the agent dynamics are implemented independently, though their behaviour still depends on the current state of the other. In my simulations I set ten agent timesteps for each global timestep, as is conventional for iDynoMiCS [83]. This is because the solution of the solute equation is computationally demanding and so the agent configuration must have changed sufficiently before the calculation of a new steady-state nutrient field is warranted.

**Algorithm 1** This algorithm shows the steps which occur when the iDynoMiCS software is run.

1. Initialise individual agents
2. Solve solute mass balances in the computational domain for the given agent distribution and bulk solute concentrations; this sets the solute concentration fields
3. While global time step < total simulation time :
   (a) While agent timestep < global timestep:
      i. Compute growth and division of agents to update agent size and mass
      ii. Apply shoving algorithm to update agent locations
   (b) Update the solute concentration fields based on new agent distributions.

### Using iDynoMiCS

As I have mentioned above, the in practice use of iDynoMiCS involves setting the values of a number of parameters at the start of the simulation using an input file. I chose parameters which represent *Pseudomonas aeruginosa* bacteria in a flow cell set up. I present those values and their literature sources in Table 3.1. Many of these values come from empirical experiments. For example, the biomass density $\rho$ and maximum specific growth rate $\mu_{max}$ of *Pseudomonas aeruginosa* are relatively well known empirically [66, 89–93]. Our choice of the solute as oxygen is also based on experiments in Kragh *et. al.* which suggest that oxygen is the limiting nutrient for *Pseudomonas aeruginosa* growth in a flow cell [91]. However, a number of my parameters used for the cell interactions are assumed. This is because these would be ether extremely difficult to measure in practice, or because they are properties of the model rather than being real physical properties (eg $k_{Shov}$). The positions of the starting cells can be specified as a full configuration, or a number of cells can be specified and iDynoMiCS will randomly distribute them across the base of the simulation. The solute concentration fields do not need to be specified in the input file because the steady-state solute concentrations are calculated from the boundary conditions and distribution of agents with their reaction rates.

In addition to specifying information about the agents and the solutes, it is

| Parameter | Values | Description | References |
|---|---|---|---|
| $S_{bulk}$ | 6.6 x $10^{-3}$ g liter$^{-1}$ | Bulk concentration of limiting growth resource (oxygen) | Saturation concentration of water at $37^oC$ [94] |
| $Y$ | 0.64 $g \cdot g^{-1}$ | Yield - grams of biomass produced per gram of oxygen consumed | [90] |
| $\mu_{max}$ | 0.29 | Maximum specific growth rate | [90] |
| $k_S$ | $8.12 \times 10^{-4}$ g liter$^{-1}$ | Concentration of Oxygen at which the growth is half maximal | [91] |
| $D$ | $2.3 \times 10^{-4} m^2 \text{day}^{-1}$ | Solute diffusion coefficient | [95] |
| Biofilm Diffusivity | 0.8 | Factor multiplying $D$ inside the biofilm | [95, 96] |
| $h$ | $80\mu m$ | Diffusion boundary layer height | [66, 97, 98] |
| $\rho$ | 200 g liter$^{-1}$ | Biomass density of bacteria | [66, 89] |
| $r_{div}$ | 2 $\mu m$ | Average cell maximum (division) radius | [90] |
| $k_{Shov}$ | 1.15 | Factor multiplying $r_{div}$ to give the shove radius | Default iDynoMiCS value [83] |
| $L_y$ | 1032 $\mu m$ | Simulation width | - |
| $N_0$ | 300 | Number of initialised cells | - |

Table 3.1: Table of the values used in my iDynoMiCS simulations.

also necessary to specify some of the variables relevant to how the iDynoMiCS algorithms are implemented and how frequently the output files are stored. I specify ten iterations of the agent time steps for each global time step, as is conventional in the use of iDynoMiCS. Once the initial values have been set in the input file, iDyniMiCS can be run from the terminal with a single line of code. iDynoMiCS uses Java to implement Algorithm 1. Output files are saved after a number of global time steps specified by the user. The output files detail the position, radius, biomass, growth rate and the genetic information for all the individual agents, and provide the nutrient field as a grid.

## 3.3 Summary

In this section, I outlined in detail how the open-source biofilm modelling software iDynoMiCS works as well as the relevant information about how I used it. I have outlined the rules which growth, division and interaction individual agents which represent the bacteria. I also outlined the equations which govern the behaviour of the solute, and detailed how the agents and the solute interact with each other over the course of the simulation.

# Chapter 4

# Simulation Methods B: Modifying iDynoMiCS to run long timescale simulations

## 4.1 Outline

In the last chapter I described in detail how the iDynoMiCS biofilm modelling software works. We saw that this was a well-established open-source software which models bacteria as individual agents coupled to a continuous nutrient field in a flow cell-like set up. In this chapter, I will outline additional simulation methods which I developed to work in tandem with the iDynoMiCS software as well as other changes that I made such as bug fixes to the iDynoMiCS software.

As I will cover in detail in the following chapters, in my research I needed to simulate biofilms over long time scales. When iDynoMiCS is run in its standard form to very long times, the number of cells becomes infeasible and the simulations too slow to proceed with. I therefore develop an algorithm which periodically removes inactive cells far below the growing front, such that a computationally feasible number of cells remain in the simulation space. The concept for this algorithm and some initial code was developed by my collaborator Gavin Melaugh. However, significant further work was done by myself in adapting the code to work on biofilms of all shapes, confirming that the removal of inactive cells was not perturbing biofilm growth and producing additional methods for

ensuring that genetic information can be recovered correctly at these long time scales.

In this chapter, I begin by outlining the basic long time simulation algorithm and the steps I took to confirm it was working as expected. I then describe the methods necessary for storing genetic information at long time scales. I finally describe big fixes to the main iDynoMiCS software and data storage methods which are also necessary for me to simulate long time scales.

## 4.2 Long Time Scales Algorithm

### Algorithm Outline

As I mentioned above, for long time simulations with iDynoMiCS I developed an additional piece of code which periodically removes inactive cells far below the growing front, such that a computationally feasible number of cells remain in the simulation space. This is achieved by pausing the iDynoMiCS simulation and removing the relevant cells, or 'clipping', and then restarting the simulation. This clipping procedure is done at regular time intervals, such that for each complete biofilm simulation there are $N$ simulation segments occurring at time intervals $T_s$, producing a total simulation time of $T = NT_s$. In practice, this involves using a python script which implements Algorithm 2 by running many short simulations with the output biofilm configuration of one simulation segment being clipped and then used as the starting configuration of the next simulation segment.

As we can see in Algorithm 2, the cells which are removed from the simulation are below a certain threshold height, a height $H$ above the base of the biofilm, which leaves the interface and the active cells intact. The threshold height is determined by calculating the position of the lowest active cell (with growth rate greater than zero) and also calculating the lowest point of the biofilm surface (which may have zero growth rate). The threshold height is then set as being the lowest of these two points. As we can see in the example configurations in Figure 4.1, this means that in the case of a smooth biofilm the threshold height is determined by the position of the lowest active cell, while in the case of a rough biofilm the threshold height is determined by the minimum interface position. Together these conditions mean the the clipping procedure does not perturb either the active layer or the position of the interface. I also subtract a buffer of $50\mu m$ from

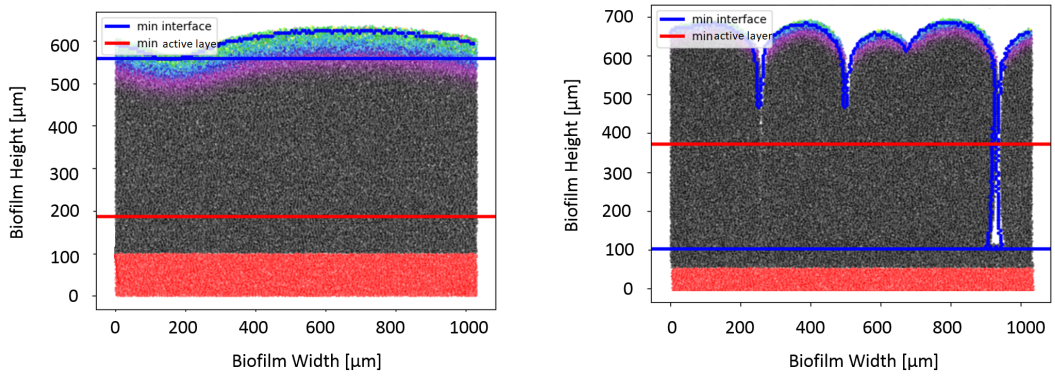
**Algorithm 2** The Long Time Simulation Algorithm, or Clipping Procedure. This procedure is used in combination with the iDynoMiCS software to reach long simulation times.

1. Run the iDynoMiCS simulation from the starting configuration up to segment time $T_s$
2. Calculate the threshold height:
   (a) Calculate the minimum interface height and the minimum of the growing layer
   (b) Set the threshold height as the lowest of the minimum interface height and the minimum of the growing layer
   (c) Subtract a buffer of $50\mu m$ from the threshold height
   (d) If the threshold height is greater than 100 $\mu m$, set it to 100 $\mu m$.
3. Remove cells below the threshold height to create the clipped cell configuration
4. Restart the simulation using the clipped cell configuration, run up to time $T_s$
5. Repeat steps 2-5 $N$ times until the end time $T = NT_s$ is reached

the calculated threshold height $H$. I set a maximum threshold height of 100 $\mu m$ to avoid large changes to the biofilm configuration in a single clipping event. In practice, the clipping procedure does not remove cells between every simulation segment, especially during early times.

When the clipped configuration is used to restart the simulation, the nutrient concentration field remains the same relative to the biofilm configuration. This is because the diffusion equation is solved at the very beginning of every simulation start, as we saw in Algorithm 1, and the diffusion boundary layer is always set a fixed height above the top of the biofilm. This means that removing inactive cells from the bottom of the biofilm configuration where nutrients does not penetrate does not change the nutrient concentration each cell experiences before and after clipping.

Producing simulations in this way means care must be taken when analysing the cumulative properties of the biofilm. For example, the total cell number of the biofilm simulation is the total of the cells in the final simulation segment plus all the cells which have been clipped out of the simulation at various points, or




(a) (b)

Figure 4.1: Snapshots of the biofilm configuration at the point of clipping. The horizontal blue line is at the height of the minimum interface height. The horizontal red line is at the height of the lowest active cell. The red cells were the cells removed. Snapshot (a) is an example where the minimum active cell determines the threshold height, and snapshot (b) is an example where the minimum interface height determines the threshold height.

in other words the effective cell number. Other cumulative properties include time and the biofilm height. Throughout the rest of this thesis, when I report these properties I report the effective property adjusted for any clipped cells. In contrast, there are some properties of the biofilm which are not cumulative and not need to be adjusted after the use of the long simulation time algorithm, such as the interface position and the thickness of the active layer of cells at the top of the biofilm.

### Testing the Algorithm

To confirm that this long time algorithm is not perturbing biofilm growth, I consider a number of test cases. Firstly, I compare the clipped simulations with a simulation without clipping, which is possible for short time scales. For starting configurations which produce biofilms with the full range of roughness behaviours which I consider later in my work, I run simulations with the same starting configuration but different simulation segment lengths $T_s$ of 0.5, 1, 2 and 4 days, along with a simulation with no clipping. To compare the results, I examined the trajectories of the active layer thickness and the interface roughness, whose calculation methods I outline in detail in the next chapter. We can see in Figure 4.2 that the trajectories of the active layer thickness and the interface roughness

are very similar. As I stated earlier, iDynoMiCS simulations are inherently stochastic, and so these trajectories are not exactly equivalent. The trajectories are close enough that we can safely assume that the clipping procedure returns simulations which are the same as without clipping.

Secondly, I also consider what happens at late times, once the steady state or relaxed behaviour of the system has been reached. At these late times, the biofilm is sufficiently large that it is not possible to run a simulation without clipping, but we can nevertheless ascertain if the clipping is perturbing the simulations by changing the clipping frequency and seeing if the trajectories remain the same. I choose three very different parameter sets, which will feature again in later chapters to repeat part of the simulation for different simulation segment times i.e different clipping frequencies. In Figure 4.3, the full trajectory is shown in black ($T_s = 2$ days) and the repeated parts of the simulation are shown in red ($T_s = 0.5$ days), blue ($T_s = 1$ day) and green ($T_s = 3$ days). In a similar manner to above, we can see that these trajectories are close enough that we can assume the clipping procedure is not significantly perturbing the simulations.

After confirming that the clipping code is not perturbing the biofilm growth, we can consider what the most optimal clipping frequency might be for simulation speed. The clipping needs to be frequent enough that the number of cells in the simulation doesn't slow the simulation down. It also cannot be too frequent, such that the simulation is being started and stopped but no cells are removed, as the starting and stopping has its own time overhead. With these considerations in mind, I chose 2 days as our standard clipping frequency.

## 4.3 Long Time Scales Algorithm with Genealogy

The long time scales algorithm needs some additional adapting to work with the genetic tracking variables described in the previous chapter, because the genealogy numbers in particular get extremely large. We recall that the cell's position in the genetic tree is uniquely identified by the family number, generation, and genealogy, seen again in Figure 4.4. We saw that a unique family number $F$ is assigned as 1...n for every initialised cell, with the progeny of these cells maintaining its family number. The generation $D$ is the number of divisions since the initialised cell. As a cell divides, one cell retains the genealogy of the mother cell $G_{mother}$, while the other gets a new genealogy $G_{new}$,

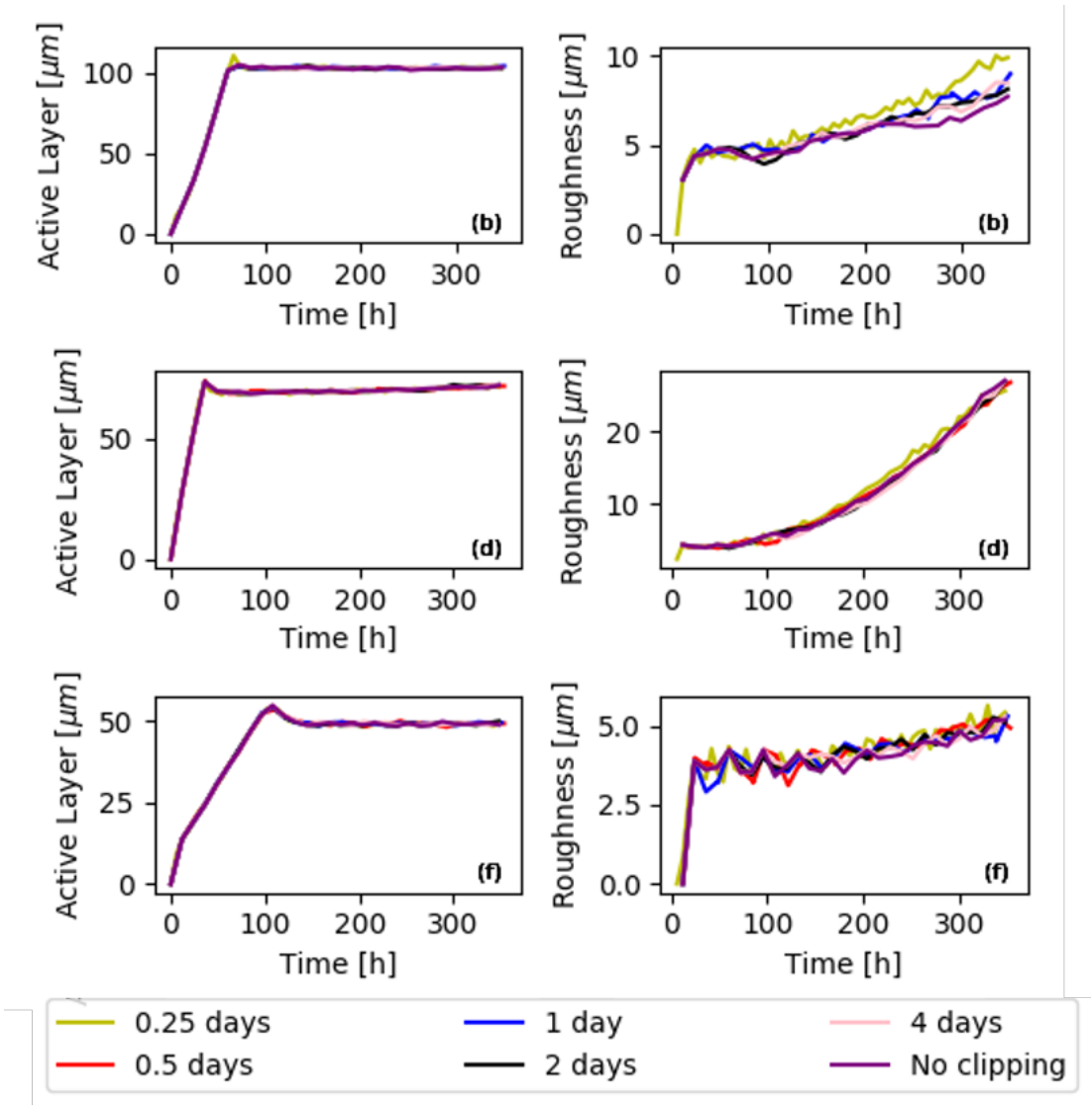



Figure 4.2: Long time algorithm test simulations. Trajectories of the active layer thickness and interface roughness for test simulations with different frequencies of clipping and without clipping. Each pair of plots (a) and (b); (c) and (d); (e) and (f) are for sets of simulations with different starting parameters. The different coloured lines on these plots are for simulations with a different simulation segment frequency $T_s$ (i.e. a different clipping frequency) and for a continuous simulation in a single segment (i.e. no clipping).

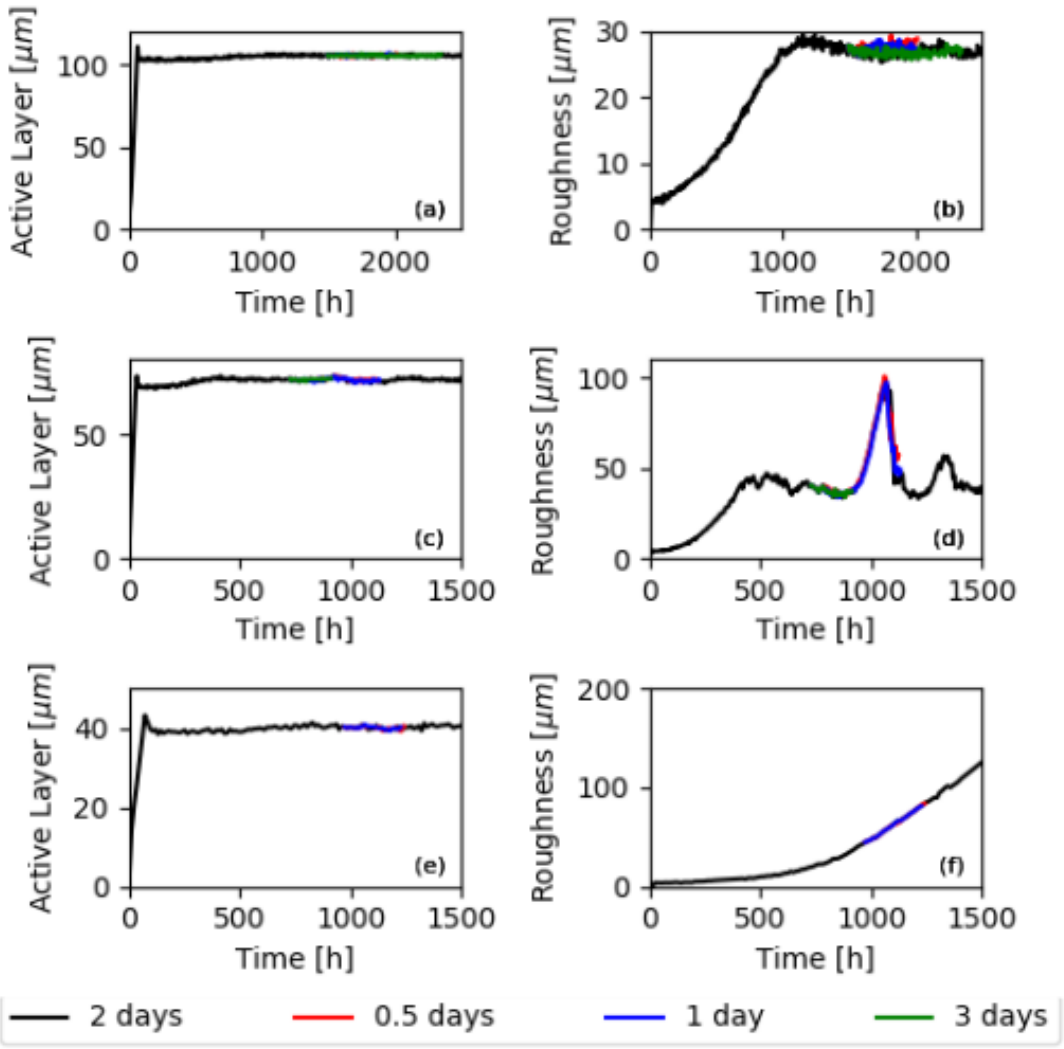



Figure 4.3: Long time algorithm test simulations. Trajectories of the active layer thickness and interface roughness for test simulations with different frequencies of clipping and without clipping. Each pair of plots (a) and (b); (c) and (d); (e) and (f) are for sets of simulations with different starting parameters. The different coloured lines on these plots are for simulations with a different simulation segment time $T_s$ (i.e. a different clipping frequency) and for a continuous simulation in a single segment (i.e. no clipping). The black line is a complete simulation with 2 days for each simulation segment, while the other segment frequencies are for a subsection of this simulation.

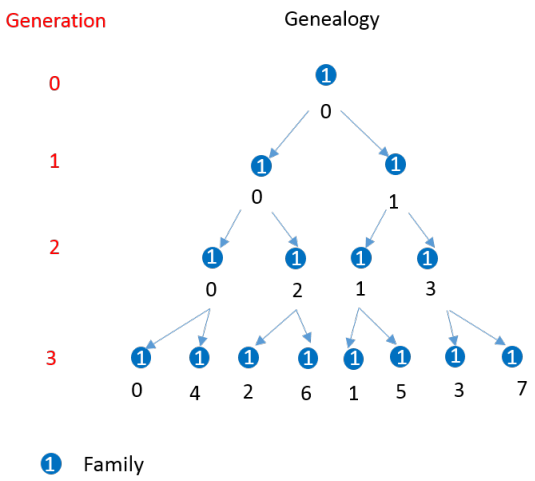



Figure 4.4: Repeat of Figure 3.3. This Figure shows a sample genetic tree beginning from a single cell. The generations of the cells are shown on the left in red. The genealogy of each cell is shown in black below each cell. The family number of each cell is shown in white on top of the cell.

as $G_{new} = 2^{D_{new}-1} + G_{mother}$, where $D$ is the generation of the newly divided cells. In practice this means the genealogy quickly becomes a very large number, reaching the variable size limit which can be stored, and even before this point is reached, tracking the genealogy slows the simulations significantly. Therefore a modification of the long time simulations algorithm is needed to be able to keep storing the cells genetic information.

To give a brief overview, the long time simulation algorithm with the genetic information works by relabelling the family, generation, and genealogy to their initialised values between each simulation segment. I also store a conversion key, which keeps track of how the labels are changed at each relabelling event. This means that I can store and process smaller numbers while still retaining the same amount of information about the genetic position of the cells and reaching larger simulation timescales. However, when analysing the outputs of these simulations, the genetic information has to be converted to the original labels ('unravelling'). In order to avoid a return to processing extremely large numbers, our analysis only involves partially unravelling these values over short time scales. However, in what follows here, I present the complete conversion to the master values which we use to check the relabelling is working. I discuss the partial unravelling methods where it is used in Chapter 8.

### Genetic variable relabelling

To begin, I outline the process of relabelling in further detail. In Figure 4.5 we can see a small snapshot of a simulation, where we can see both the 'master' genetic tree on the left hand side, and the equivalent relabelled genetic tree on the right hand side. Each time a new simulation segment is produced, the generation and genealogy of each cell is set to zero, and each cell is assigned a unique family number. This means the genetic parameters of the cells at the start of each new simulation segment looks like a configuration of cells that could be used to initialise a whole simulation, with the family number uniquely identifying each cell as the generations and genealogies are set to zero. By the end of each simulation segment, there are many cells with the same family number, but the cells are still uniquely identifiable with a combination of their family, generation and genealogy number, in the same way as cells at the end of a single simulation without any clipping or relabelling would be.

As I mentioned earlier, an important part of the relabelling process is storing the conversion keys to ensure that the same amount of genetic information about the cells is retained. For each relabelling event, I store a conversion key for that particular event, as we see in Figure 4.5. After a relabelling event, the family number is what uniquely identifies the cell, as the generation and the family are set to zero. This means that the conversion key between two segments, for example between segments 2 and 3, links the family number at the beginning of segment 3 to the family, generation and genealogy number of the same cell at the end of segment 2. Since there are many simulation segments over the course of a complete simulation, cells are relabelled many times over the course of a simulation. This means that many conversion keys are used over the course of a simulation. Each conversion key allows us to connect the genetic labels for a given segment to those of the previous segment, rather than to the master labels of that cell.

### Genetic variables unravelling

As I mentioned, I describe here how to return these relabelled values to their master values, which I use for demonstrating that this procedure retains the complete genetic information about all the cells in the biofilm. The complete unravelling of the genetic information for a single cell is done according to

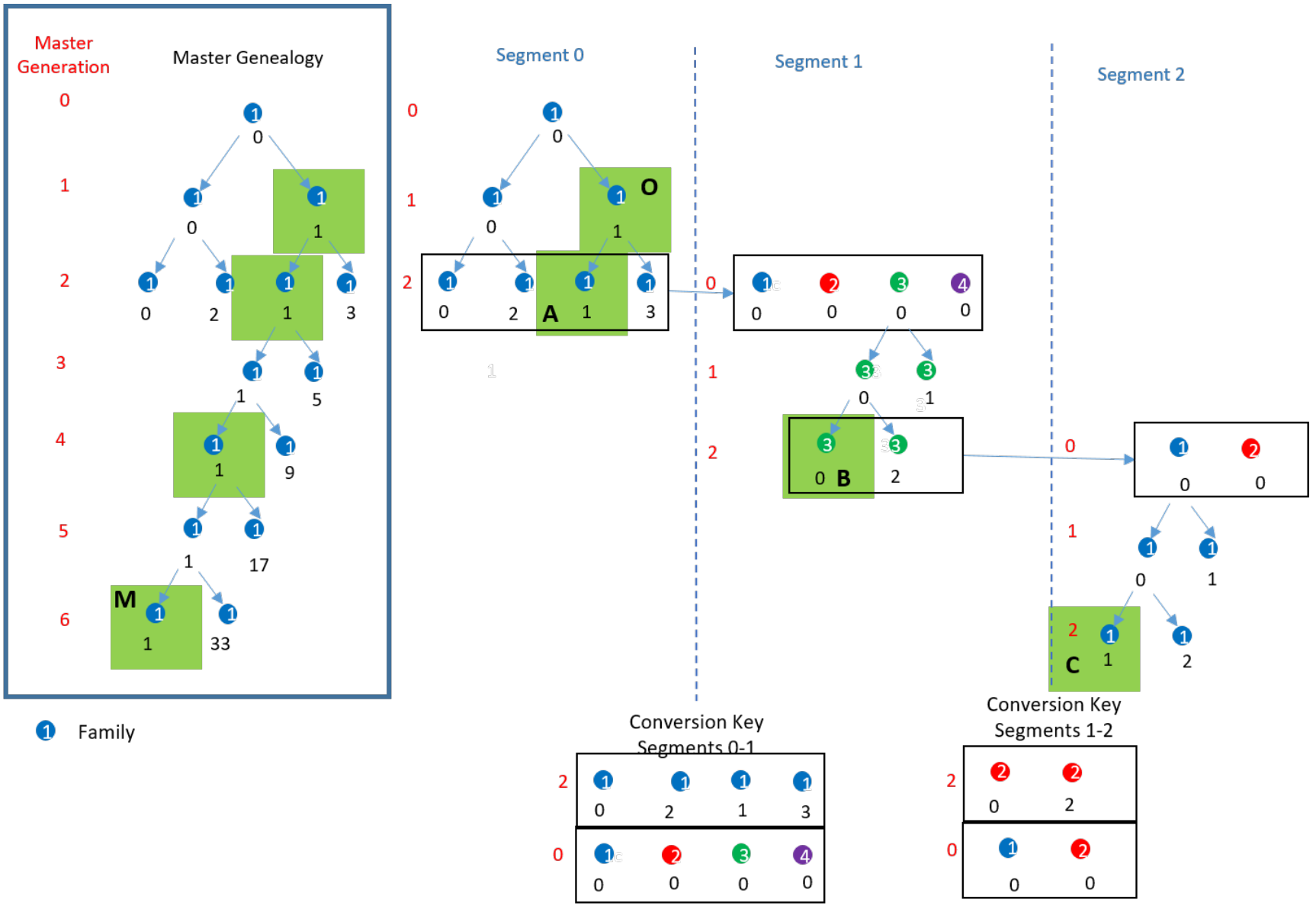



Figure 4.5: Diagram of how the cells genetic information is relabelled during the long time simulation algorithm. In box on the left is the master genetic tree. On the right hand side is the same genetic tree that this time has been relabelled over the course of three simulation segments. Below each segment division is a partial conversion key for that segment transition. Family numbers are shown in white on top of the cell. Generation numbers are shown in red to the left hand side. Genealogy numbers are shown in black.

**Algorithm 3** Algorithm for converting the relabelled genetic variables to their master values, for a single cell at one time point. The cell being considered is in segment $S$ and has family number $F_S$, generation $D_S$ and genealogy $G_S$. Each of the genetic variables for in intermediate simulation segments are denoted with the subscript $s$. Master genetic variables are denoted with subscript $M$. The conversion key between simulation segments $s-1$ and $s$ is denoted $K_{s-1,s}$. Genealogy in binary format is denoted with superscript $B$, and the function for combining two genealogies in binary format, which is described further in the text, is $H$.

1. For each segment s, beginning with the segment $S$ the cell is in and going backwards through all the segments to the initial segment 0:
   (a) Set $F_s$, $G_s$ ad $D_s$ (from S at the beginning or from the $s-1$ that is found along the way.)
   (b) Find previous segment values using conversion key: $K_{s-1,s}[F_s] = (F_{s-1}, D_{s-1}, G_{s-1})$
   (c) Increment Generation: $D_M = D_M + D_{s-1}$
   (d) Increment Family: $F_M = F_{s-1}$
   (e) Increment Genealogy:
      i. Convert genealogy to binary coding: $G_{s-1} \rightarrow G^B_{s-1}$ using $D_{s-1}$
      ii. $G^B_M = H[G^B_M, G^B_{s-1}]$
2. Return genealogy to original coding: $G^B_M \rightarrow G_M$

Algorithm 3. This is a non-trivial process for two main reasons. Firstly, each conversion key only relates the current segment to the previous one. This means many relabelling events must be unravelled for each cell to reach the master genetic information. It also means the genetic information must also be unravelled in the reverse order to which they were relabelled, beginning with the segment number of the cell we are unravelling $S$ and going through each segment to the initial simulation (zeroth) segment. Secondly, it is necessary to unravel all three variables alongside each other, because all three variables are needed to uniquely identify the cell and therefore correctly use the conversion key. I describe the process for unravelling the family, generation and genealogy one by one for clarity, though the full algorithm unravels them simultaneously as in Algorithm 3.

I begin by describing the process for the family number. Since the family number is inherited by each of the progeny of a cell, finding the master family $F_M$ number means finding the family number of the cell when it was initialised ($F_O$) either

at the start of the simulation or when it was produced after a division event). We see in Algorithm 3 that to find $F_M$ it is necessary to begin with the current segment S, and to find the family number of our chosen cell's ancestors in the previous segment. I repeat this process, working backwards, for each segment until segment 0 is reached i.e. $F_S \rightarrow F_{S-1}$; $F_{S-1} \rightarrow F_{S-2}$, . . . .$F_O$. As an example of how this works, we consider Figure 4.5, where the green highlighted cell O has undergone relabelling both over the course of division and using the relabelling algorithm to become green highlighted cell C. We see that cell C has family number 1. To find the family number of the cells ancestors in segment 1, we look at the conversion key for segment 1-2, and we see that cells with family 1 in segment 2 were labelled family 3 in segment 1. By looking at the conversation key for segments 0-1, we see the ancestors of family 3 in segment 1 were family 1 in segment 0. Therefore, the master family is $F_0 = 1 = F_M$.

Next, I consider how the generation is unravelled. To convert a relabelled cell generation number to the master generation number $D_M$, we need the total number of generations since the cell was initialised. This can be done by counting the number of generations which have passed in each segment i.e $D^i_M = D^j_S + D^k_{s-1} + ... + D^i_M$. This is simple in principle, but we need to make sure we are counting divisions which occurred for that particular cell, which requires the use of the conversion key and therefore the family number. We can see how this works in practice by again considering Figure 4.5. We can see that 2 generations have passed in segment 2 in the lineage of cell C, as cell C is in generation 2 of its current segment 2 ($D^C_2$). By using the conversion key for this segment 1-2, we see that cell B has undergone 2 divisions in segment 1 ($D^B_1$). By the same reasoning, we see cell A has undergone 2 divisions in segment 0 ($D^A_0$). Added together, this means the master generation of cell C is $D^D_M = D^A_0 + D^B_1 + D^C_2 = 2 + 2 + 2 = 6$.

Finally, we consider how the relabelled genealogy is converted to its master value. This is a rather more involved process than unravelling the family or generation. Though there is likely more than one way of doing this, in my work I have chosen to convert the genealogy value into binary format. In Figure 4.4 we can see an example genetic tree with the usual representation of the genetic variables displayed alongside the equivalent genetic tree with the genealogy in binary representation. There are two things to notice about this representation. Firstly, when a cell divides in binary representation, one of the daughter cells gets a zero in front of the mother's genealogy number, and the other gets a 1

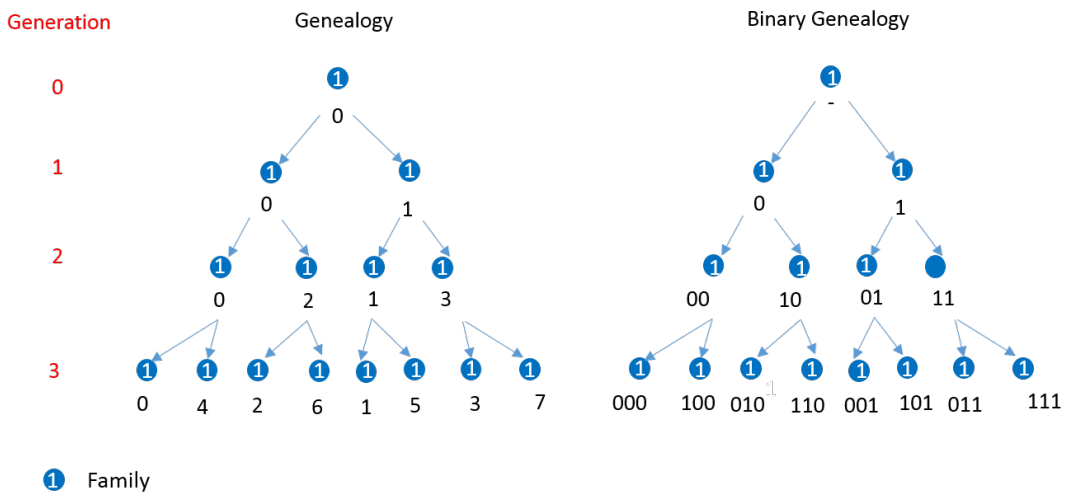



Figure 4.6: This Figure shows a sample genetic tree beginning from a single cell with the genealogy labelled in two different formats. The genealogy of each cell is shown in black below each cell. The genetic tree on the left has the genealogy in the original format, and the genetic tree on the right has the genealogy labelled in binary format. The generations of the cells are shown on the left in red. The family number of each cell is shown in white on top of the cell.

e.g. when a cell with genealogy 1 divides, the daughter cells have genealogies of 01 and 11. Secondly, because of this, the digit length of the binary genealogy is the same as the generation e.g. genealogy 110 has a generation of 3. This means that the binary genealogy now contains information on both the generation and genealogy, so we only have to deal with the family and the binary genealogy rather than all three genetic variables.

Using the genealogy in binary format allows us to more easily find the master genealogy $G_M$. We now consider Figure 4.7, which is the same as Figure 4.7 except the genealogy is in binary format. Finding the master genealogy involves determining how much the genealogy numbers have changed over the course of each segment and combining this information. Using the binary genealogy format and the conversion keys as before, we can collect the genealogies of cell C (00) , B (00) and A (01) at the points of relabelling, and place them one in front of each other. We then have the same genealogy cell M (000001), i.e the master genealogy. This is a result of the binary format genealogy for two daughter cells involving putting the 0 or 1 in front of the mothers genealogy as we saw previously, but we are collating the additional 0s and 1s which have accumulated over the course of an entire simulation segment.

When unravelling the genealogy in practice, we are of course not given the

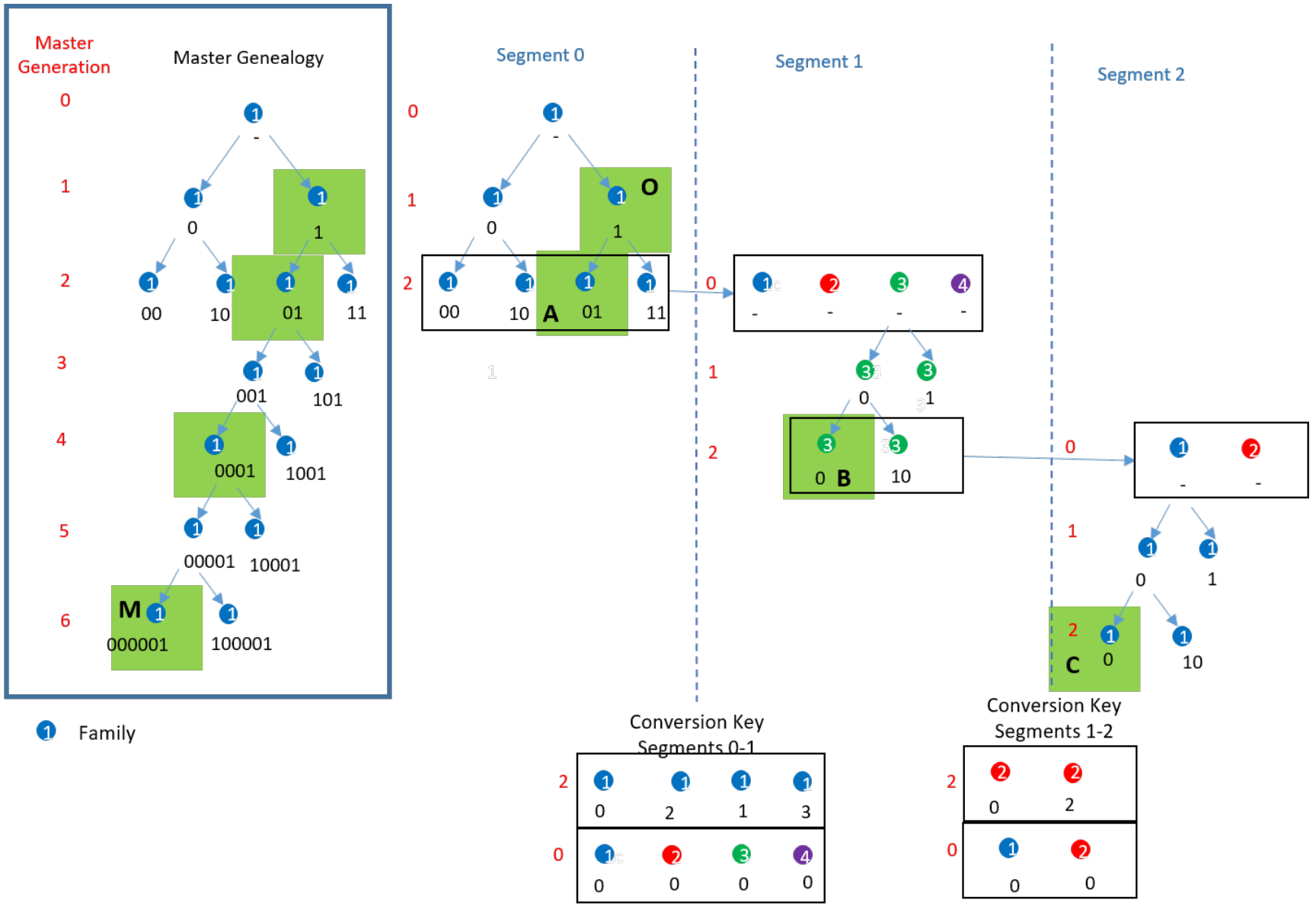



Figure 4.7: Diagram of how the cells genetic information is relabelled during the long time simulation algorithm, this time with the genealogy in binary format. In box on the left is the master genetic tree. On the right hand side is the same genetic tree that this time has been relabelled over the course of three simulation segments. Below each segment division is a partial conversion key for that segment transition. Family numbers are shown in white on top of the cell. Generation numbers are shown in red to the left hand side. Genealogy numbers are shown in black in binary format.

whole genetic tree in binary format so we must first convert the genealogy into binary format, as we see in Algorithm 3. As I mentioned, this involves using the generation, because we can see in Figure 4.6 there are many identical genealogies within one family in the original format, only distinguished by their generation. As an example, consider a cell in generation 4 with a genealogy number of 2. I firstly convert this genealogy into its binary format, i.e. $2 \rightarrow 10$. Secondly, I add zeros in front of this to make the value the same length as the generation, i.e. $10 \rightarrow 0010$, allowing it to be distinguished from other cells with the same genealogy at different positions in the tree. Once the genealogies are converted to binary format, it is possible to place them one in front of each other to obtain the master genealogy in binary format, and finally convert back to the original format to find the master genealogy $G_M$.

I have now outlined the complete process via which I relabel the genetic parameters, and how they can be converted to their master values. To test these procedures, I run simulations with and without relabelling for short simulation times to confirm the same genetic information is retained. As we discussed, in the chapters which follow I use this relabelling algorithm to reach long simulation times whilst still storing the complete genetic information. When I analyse the genetic information in detail in Chapter 8, I do not convert the genetic variables to their master values because we would return to the problem of having to do calculations with very large numbers. Instead, we partially unravel the genetic information and use other shortcuts, which we detail in the relevant section of Chapter 8.

## 4.4 Additional Technical Aspects of the Simulations

In addition to the long time simulation algorithms, to perform large simulations on the scale my work required I also had to use high performance data storage methods to make the analysis of such large sets of data tractable. We used the HDF file library which is specifically designed for large data sets, and converted all the of the iDynoMiCS outputs into this format before doing the further analysis. While this was a lengthy process, it only needed to be done once and allowed all further manipulation of the data to be done significantly more quickly.

During my research, I found a number of bugs in iDynoMiCS that needed to be fixed in order to proceed with my work. Firstly, I discovered that iDynoMiCS was not reading in the starting cell configuration correctly, which had a particular impact on my simulations because I was creating many simulation segments which involved restarting the simulation many times. A second bug in iDynoMiCS became obvious when I began investigating the relatedness of cells in the biofilm - it was not possible to identify each of the cells uniquely. This led me to discover that the calculation of the new genealogies after a cell had divided was not implemented correctly. These bugs were both fixed with the help of Professor Jan Kreft (University of Birmingham), one of the original developers of iDynoMiCS. Both of these fixes have now been added to the open source version of the software.

## 4.5 Summary

In this chapter I outlined the simulation methods I developed to use in conjunction with the open source software iDynoMiCS, which we outlined in the previous chapter. I firstly outlined a long time simulation algorithm, which periodically removed cells from the simulation which were no longer growing because they were far below the interface. I secondly outlined the specific additions to this algorithm which needed to be made in order to store the complete genetic information of the biofilm over these long simulation times, as the genealogy variable is larger than can be stored for large biofilm sizes. Finally, we detailed the additional data storage methods and iDynoMiCS bug fixes which were necessary in order for both of these algorithms to work effectively.

I use these simulation methods throughout the rest of this thesis. In the next chapter, we will see the early time simulation results which motivate the large simulations which I have detailed the simulation methods here. In the rest of the results chapters (Chapters 5 - 8), I analyse the outputs of simulations using the long time simulation algorithms in combination with iDynoMiCS. As I mentioned earlier, the methods I use to analyse the outputs of these simulations will be presented in the chapters in which they are used.

# Chapter 5

# The Interface Roughness and the Active Layer

## 5.1 Outline

We have seen in Chapter 2 that there is a great interest in understanding the conditions which produce different biofilm morphologies and defining the phase behaviour of the non-equilibrium dynamics of biofilm growth. We saw that a variety of biofilm shapes, both smooth and rough, have been observed experimentally and theoretically. Though distinct phases of behaviour have been observed under different conditions, and frameworks for describing their phase behaviour have been proposed, the phase behaviour of biofilm morphology remains a topic of debate. In this chapter, I contribute to this debate by investigating using individual based-simulations the different biofilm spatial structures which are produced for different conditions. As we have described in Chapter 3, I choose to work *in silico* as simulations are quicker to run than lab experiments and it is therefore easier to explore the relevant parameter space, as a route to more directed experimental work.

In characterising the phase behaviour of biofilm growth, I choose to focus on two key quantities that have emerged from the literature. The first quantity is the interface roughness, which I define as the standard deviation of the biofilm height. Roughness is one of the most commonly used quantities in the literature to characterise the spatial structure of the biofilm [16, 21, 99]. The second quantity

is the thickness of the layer of growing layer of cells close to the interface, or as it is often known, the active layer thickness. The active layer thickness has been suggested to be a driver of the global spatial structure of the biofilm as well as the genetic structure of biofilms, and I therefore think it is a good quantity to focus on given my stated aim of better understanding the link between the genetic and spatial structure of biofilms [28, 64, 65, 75].

In this chapter I define the methods I use to calculate the interface roughness and the active layer thickness, and analyse their dynamical trajectories. My main focus here is to examine the steady-state dynamics of these key quantities, and how they relate to the input parameters of the simulation. In Section 5.3.1 I examine the active layer thickness, which reaches its steady state very rapidly as the biofilm forms. I go on in Section 5.3.2 to look at the behaviour of the interface roughness, which takes much longer to reach its steady state and requires the use of the long time simulation algorithm that I outlined in Chapter 4. At these long times I observe three distinctive regimes of roughness behavio ur, which I discuss in this chapter and investigate further in the work which follows in the rest of this thesis.

## 5.2 Analysis Methods

### 5.2.1 Interface Calculation

I begin by defining the interface. There are multiple ways to define an interface, particularly when it comes to interfaces with complex shapes. I start by discussing the simplest method - which has sometimes been called the single-valued interface method [100, 101]. This involved slicing the biofilm into vertical strips, and finding the bacterial cell with the highest position in each of the strips. As we see in Figure 5.1(a), this means that every position in the $x$ direction, there is a single interface height position in the y direction. This works effectively for smooth biofilms, but for biofilms with a complex interface structure, usually when the interface is very rough, the detail of the interface shape is lost. In Figure 5.1(a), we can see that defining the interface in this way does not capture the troughs between the biofilm fingers, meaning the interface roughness (standard deviation of the interface points) would be underestimated.

As I have stated, the interface roughness is an important quantity in our analysis,

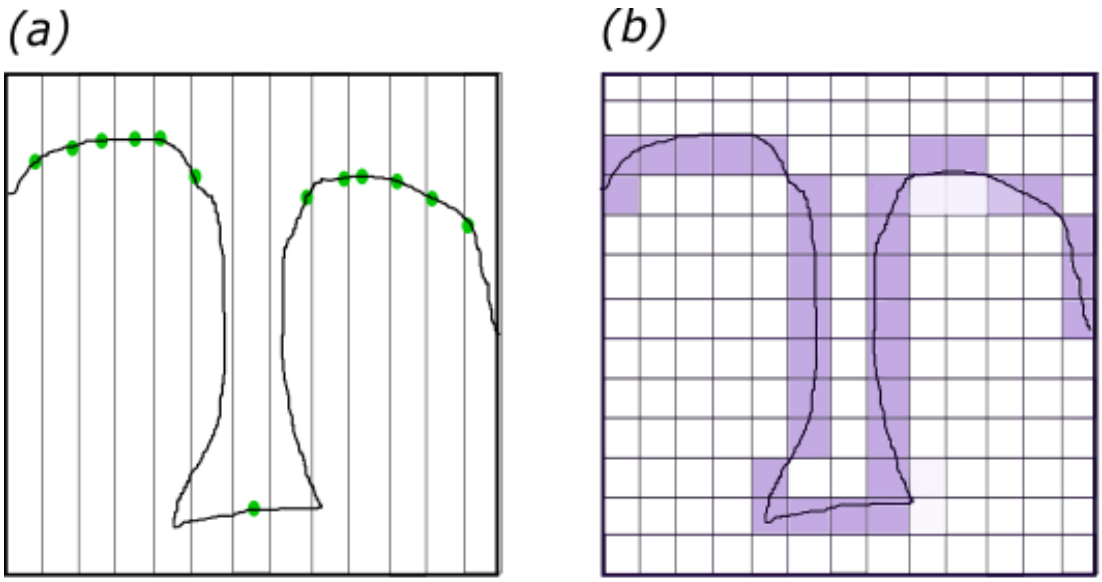



Figure 5.1: Sketches to demonstrate the difference between the single-valued interface (SVI) calculation and the multi-valued interface (MVI) calculations. (a) shows the SVI calculation, where the highest cell in each vertical bin is coloured in green. (b) shows the MVI calculation, where each grid square which contains biomass and is adjacent to a grid square not containing biomass is coloured blue.

and capturing the details of the interface correctly is essential. Therefore, I define what is sometimes known as a multi-valued interface [100, 101]. I do this by binning the bacterial biomass into a grid with $D$ horizontal and $H$ vertical slices and then searching for grid squares which contain biomass but have a nearest neighbour that does not contain biomass. This produces a set of grid squares corresponding to the interface $\{k\}, k = 1, ...N_{int}$ where $N_{int} \geq D$. As we can see in Figure 5.1, the troughs between the biofilm fingers are now more accurately captured. We can also see that defining the interface in this way means there can be an interface where there can be multiple vertical ($x$) points for each horizontal points ($y$) - hence its description as a multi-valued interface.

I also define the interface width, or the interface roughness, as the root mean square of the points on the interface, summed over the vertical direction $x$,

$$W(L,t) = \left\langle \left[h(x,t) - \langle h(x,t)\rangle_L\right]^2 \right\rangle^{1/2}. \tag{5.1}$$

Since $h(x,t)$ is a multi-valued interface, $\langle h(x,t)\rangle_L$ is defined as the mean value of $k$ vertical coordinates $h_k$ along the interface of length $L$, such that $0 < h_k < L-1$

and $1 < k < N_{int}$. Namely,

$$\langle h \rangle_L = \frac{1}{N_{int}} \sum_{k=1}^{N_{int}} h_k \tag{5.2}$$

with $N_{int}$ being the number of point along the interface, which for an mutli-valued interface is more than the number of bins in the width D ($N_{int} > D$). This means I am taking the standard deviation of the interface, including any overhangs.

### 5.2.2 Active Layer Thickness Calculation

I now define what the active layer the active layer is, before describing how I calculate the active thickness. We can see the active layer in Figure 5.2, where we can see that the actively growing cells are the ones at the top of the biofilm which have access to nutrients. We can see from the left panel that there is no nutrients inside the biofilm as it has been depleted by the growing cells. I begin by defining a threshold growth rate; cells which grow faster than this rate are defined to be part of the active layer. In contrast to this approach, I consider a cell to be in the active layer when it grows at greater than 0.1% of the specific growth rate of a cell $\mu$ that is possible under the conditions of the simulation (ie.e for given $\mu_{max}$ and $S_{bulk}$). The specific growth rate is determined by the Monod function I detailed in Equation 3.3 in Chapter 3, and so the condition for a cell to be part of the active layer is,

$$\mu > \frac{\mu_{max}}{1000} \frac{S_{bulk}}{k_s + S_{bulk}} \tag{5.3}$$

where $k_S$ is the solute concentration at which the growth is half maximal and $S_{bulk}$ is the bulk solute concentration.

Having defined the condition for a cell to be in the active layer, I now outline how the active layer thickness is calculated. As with the interface definition above, I define a grid with $D$ horizontal and $H$ vertical slices of width $8\mu m$ such that the system size width $L = D \times 8\mu m$ and the system height $S = H \times 8\mu m$. For each of the $D$ vertical columns, I total the number of grid squares whose biomass has an average specific growth rate above the threshold in Equation 5.3 in order to find the local active layer thickness. For some biofilm configurations, for example when the biofilm is rough, there can be a growing layer both at the leading edge of the biofilm and further down. In an analogous manner to the

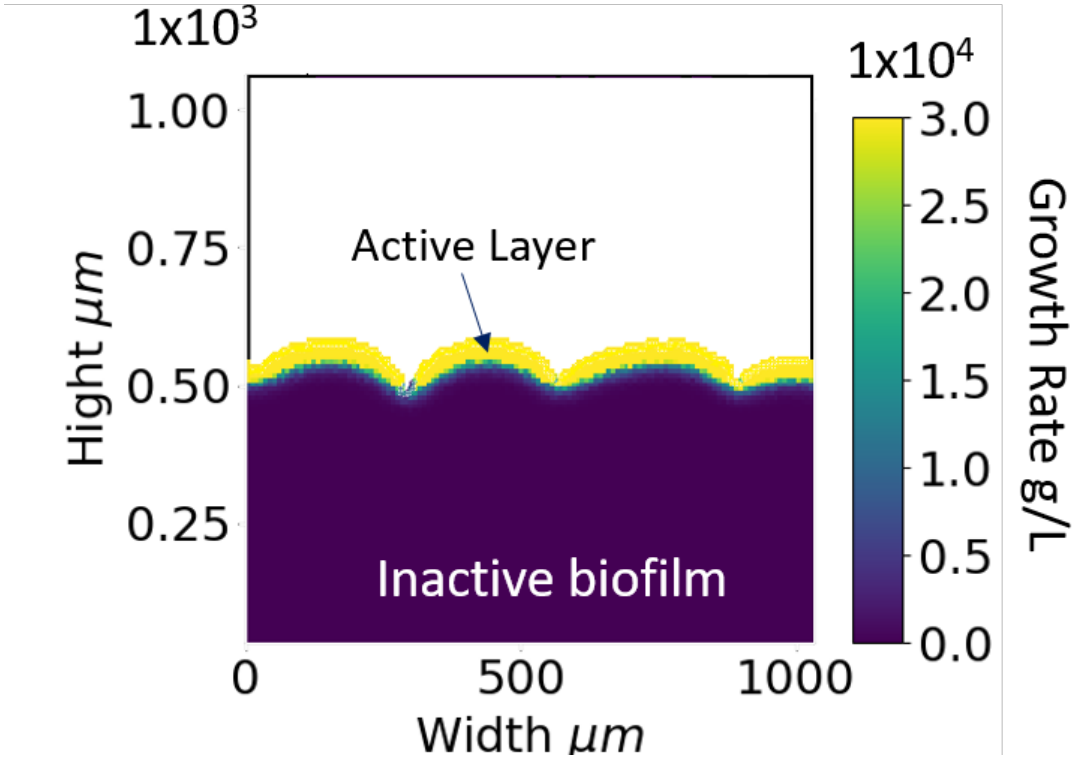



Figure 5.2: This Figure shows a snapshot of biofilm growth, where the growth rate of the cells are shown in the colour scale. We can see that there is an active layer of cells at the top of the biofilm, known as the active layer. Inside the biofilm, cells are not growing and have zero growth rate.

multi-valued interface defined above, in the case where there are active sites both at the leading edge of the biofilm and far below the leading edge within the same vertical strip, we take the active layer thickness to be the total thickness as the sum of the two parts. Once the active layer thickness for each vertical strip has been found, the mean active layer thickness (averaged over all columns) and its standard deviation can be calculated for a particular biofilm configuration.

## 5.3 Results

As I mentioned earlier, my main aim in this chapter is to investigate the behaviour of the active layer thickness and the interface roughness, for parameters that lead to different spatial configurations of the biofilm. Throughout this chapter, I will plot these quantities against the cell number, rather than time. The cell number represents the size of the biofilm, rather than the time it has been growing. As we can see in Figure 5.4, time is linearly related to the cell number for different

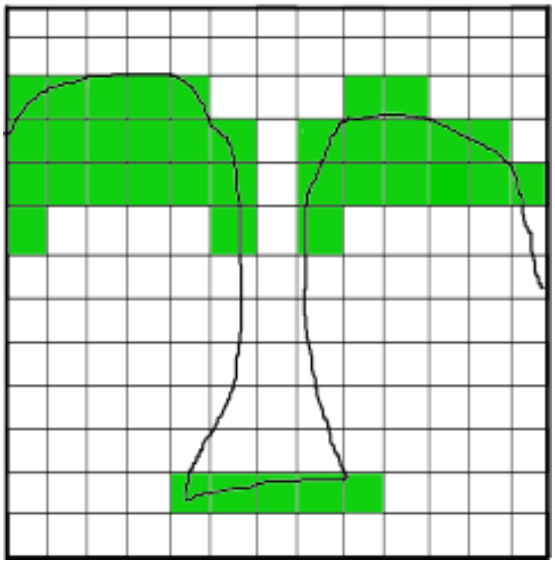


Figure 5.3: This Figure shows a sketch to demonstrate how the active layer is calculated. The grid squares which are considered part of the active layer are coloured in green.

simulation starting parameters, though the gradient is not always the same. I will discuss my reasons for doing this in more detail later in the chapter, when these reasons are more coherently explained as the relevant results have been presented.

### 5.3.1 The Active Layer Thickness

I begin by investigating the behaviour of the average active layer thickness for different simulation starting values. The key parameters that I vary are the bulk nutrient concentration $S_{bulk}$, the height of the diffusion boundary layer $h$, the maximum specific growth rate of the bacteria $\mu_{max}$, the growth substrate diffusion coefficient $D$, the yield with which cells convert substrate to biomass $Y$ and the cell biomass density $\rho$. These parameters were varied over ranges of parameters centered around the literature values detailed in the Table 3.1 in Chapter 3. Examples of these trajectories are plotted in in Figure 5.5, where each of the subplots has trajectories of simulations with one of the key parameters varied.

The most striking feature of the trajectories in Figure 5.5 is that the average active layer thickness reaches a steady state early on in the biofilm's growth, and

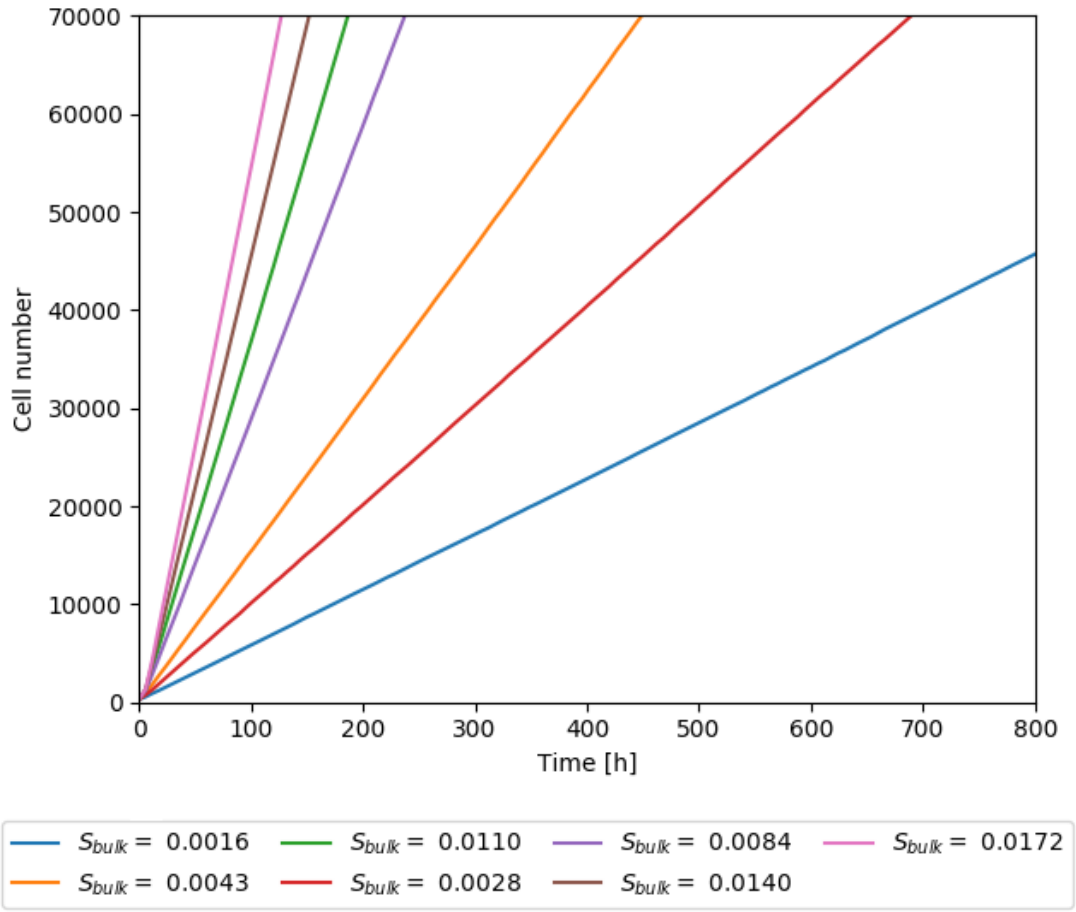



Figure 5.4: This Figure shows the linear relationship between time and cell number for simulations with varying bulk nutrient concentration $S_{bulk}$. Input parameters are otherwise as specified in Table 3.1.

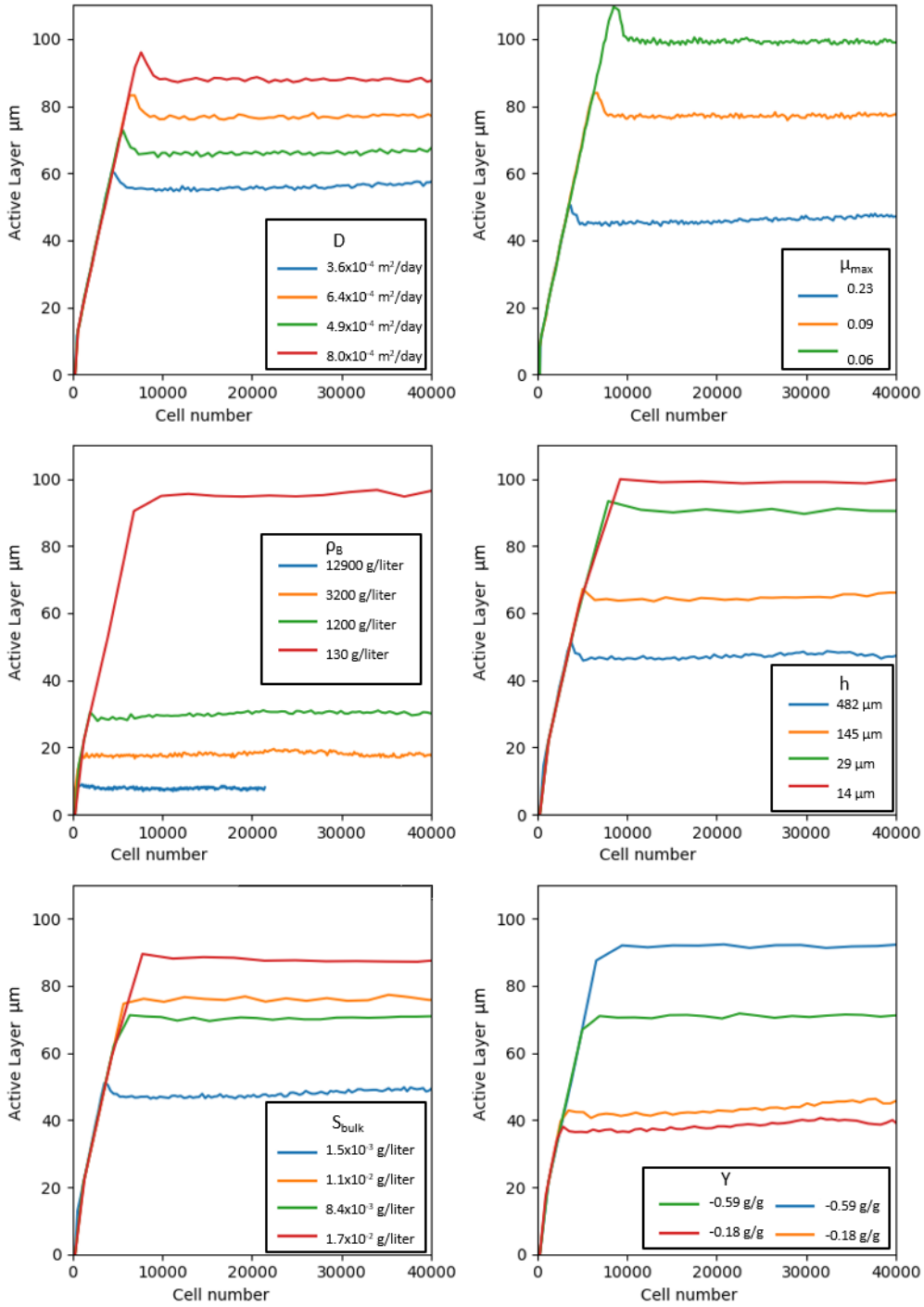



Figure 5.5: This Figure shows the active layer thickness trajectory for simulations where the diffusivity $D$, maximum specific growth rate $\mu_{max}$, biomass density $\rho$, diffusion boundary layer height $h$, yeild $Y$ and bulk nutrient concentration $S_{bulk}$ are varied in each of the subplots respectively. The input parameters for the simulations are as in Table 3.1 unless otherwise specified in the figure legends.

the trajectories follow a similar pattern for a wide range of starting parameters. In the very early stages when all cells have access to the substrate, the active layer thickness increases linearly with time. If all cells had equal access to nutrients and so were all dividing at the same rate, we might expect there to be an exponential increase in the active layer thickness at early times. However, as not all cells have equal access to the substrate, some are dividing much more slowly than others so the increase in the size of the biofilm and the active layer thickness is slower than exponential. After the biofilm has reached a critical thickness, the substrate can no longer penetrate all the way down so the active layer thickness levels off and remains constant throughout the rest of the simulation.

We can also see in Figure 5.5 that for some of the simulations, the active layer thickness is briefly thicker than the steady state active layer thickness i.e. there is a 'blip' between the early time growth and the steady state. Careful examination of images of the biofilm configurations leads me to suggest that this is because the substrate can accumulate on the base of the biofilm container. This means that when the biofilm is transitioning between all the cells growing and just some of them, the accumulated substrate can temporarily continue to supply cells at the bottom of the biofilm and of the active layer until it is used up, thus transiently increasing the thickness of the active layer. Since this behaviour depends on how far inside the biofilm nutrients is able to diffuse and many of the parameters varied will impact this in different ways, it is not unsurprising this behaviour occurs for some parameter sets and not others. Nevertheless, further investigation would be needed to establish why this has occurred for the particular parameter sets we can see in Figure 5.5.

Since the average active layer thickness becomes constant quite early on in the simulations, it is possible to investigate the relationship of this steady state active layer thickness to the input parameters of the simulation. In Figure 5.6 we can see the average active layer thickness plotted against the values of $S_{bulk}$, $\mu_{max}$, $h$, $Y$, $\rho$ and $D$ in a set of simulations where all the other starting parameters are kept the same. We can see that each of these parameters has a distinct effect on the average active layer thickness. Increasing the diffusivity $D$ increases the active layer thickness because the substrate can diffuse more rapidly down inside the biofilm, so more cells are growing at a slower rate. Increasing the bulk nutrient concentration $S_{bulk}$ increases the active layer thickness as there is a greater supply of nutrients which can travel further. The active layer thickness decreases for increasing the maximum specific growth rate $\mu_{max}$, yield $Y$ and cell

biomass density $\rho$ because these parameters all increase the amount of substrate which is consumed by the cells, and so the substrate travels less far before it is used up. Increasing the height of the diffusion boundary layer $h$ decreases the active layer thickness because nutrients have further to diffuse before reaching the biofilm.

Plotting these relationships again on a log-log plot in Figure 5.7, we can see that these are all power law relationships. My intention in doing this was to see if I could find a way to predict the average active layer thickness from the input parameters. This was inspired in particular by the work of Nadell *et. al.*, who we saw in Chapter 2 argue the active layer thickness is correlated with a non-dimensional quantity $\delta = \sqrt{S_{bulk}DY/\mu_{max}\rho h^2}$ which is a combination of input parameters. While the input parameters which make up Nadell *et. al.*'s $\delta$ parameter are the same key parameters I identify in my simulations, the exponents I measure in Figure 5.7 are not consistent with Nadell *et. al.*'s parameter. Nadell *et. al.*'s simulation set up is not the same as ours, and I consider the possible reasons for this discrepancy in the discussion later in this chapter. Further investigations may yet reveal that is it possible to calculate the average active layer thickness from the input parameters in my simulation set up.

### 5.3.2 The Interface Roughness

I now investigate the interface roughness, the second important quantity in my analysis. In Figure 5.8, we can see the trajectories of the interface roughness plotted for the same simulations as Figure 5.5, where each subplot corresponds to variation of a single parameter. In contrast to the active layer thickness, the interface roughness does not appear to consistently reach a steady state over the time scales presented, which makes it difficult to compare. In Figure 5.9, we can see the log-log of the interface roughness. This does not consistently reach a linear trajectory either, which shows that it would also be difficult to compare the rate at which the interface roughness is increasing. Since classifying both the steady state and the rate of increase of the interface roughness is not possible at these time scales, this motivates my decision to run long time scale simulations to reach the steady state interface roughness, which I detail in the next section.

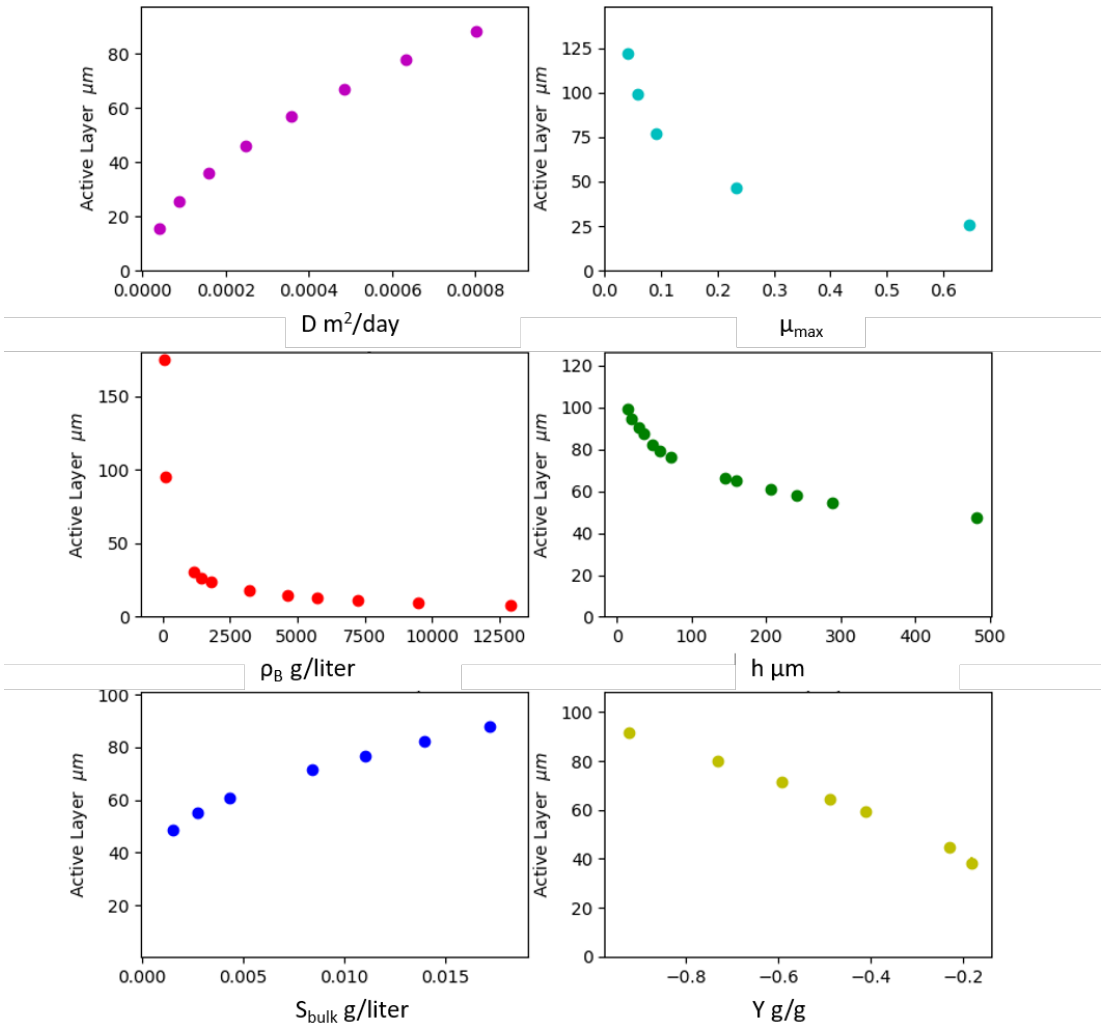



Figure 5.6: This Figure shows the steady state average active layer thickness against the input parameter which as been varied for that simulation. The diffusivity $D$, maximum specific growth rate $\mu_{max}$, biomass density $\rho$, diffusion boundary layer height $h$, yield $Y$ and bulk nutrient concentration $S_{bulk}$ are varied in each of the subplots respectively. The input parameters for the simulations other than those on the $x$-axis are as in Table 3.1.

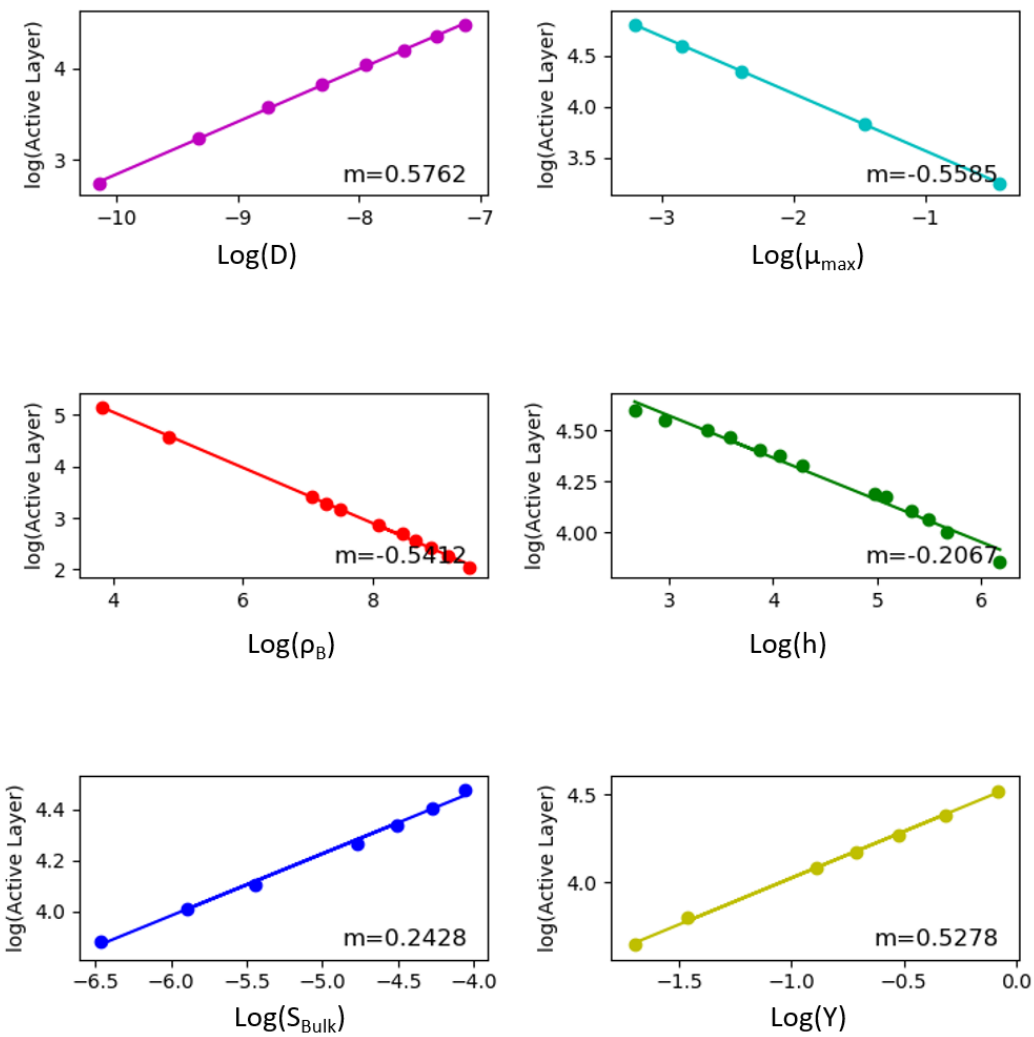



Figure 5.7: This Figure shows the log of the steady state average active layer thickness against the log of the input parameter which as been varied for that simulation. The diffusivity $D$, maximum specific growth rate $\mu_{max}$, biomass density $\rho$, diffusion boundary layer height $h$, yield $Y$ and bulk nutrient concentration $S_{bulk}$ are varied in each of the subplots respectively. The input parameters for the simulations other than those on the $x$-axis are as in Table 3.1.

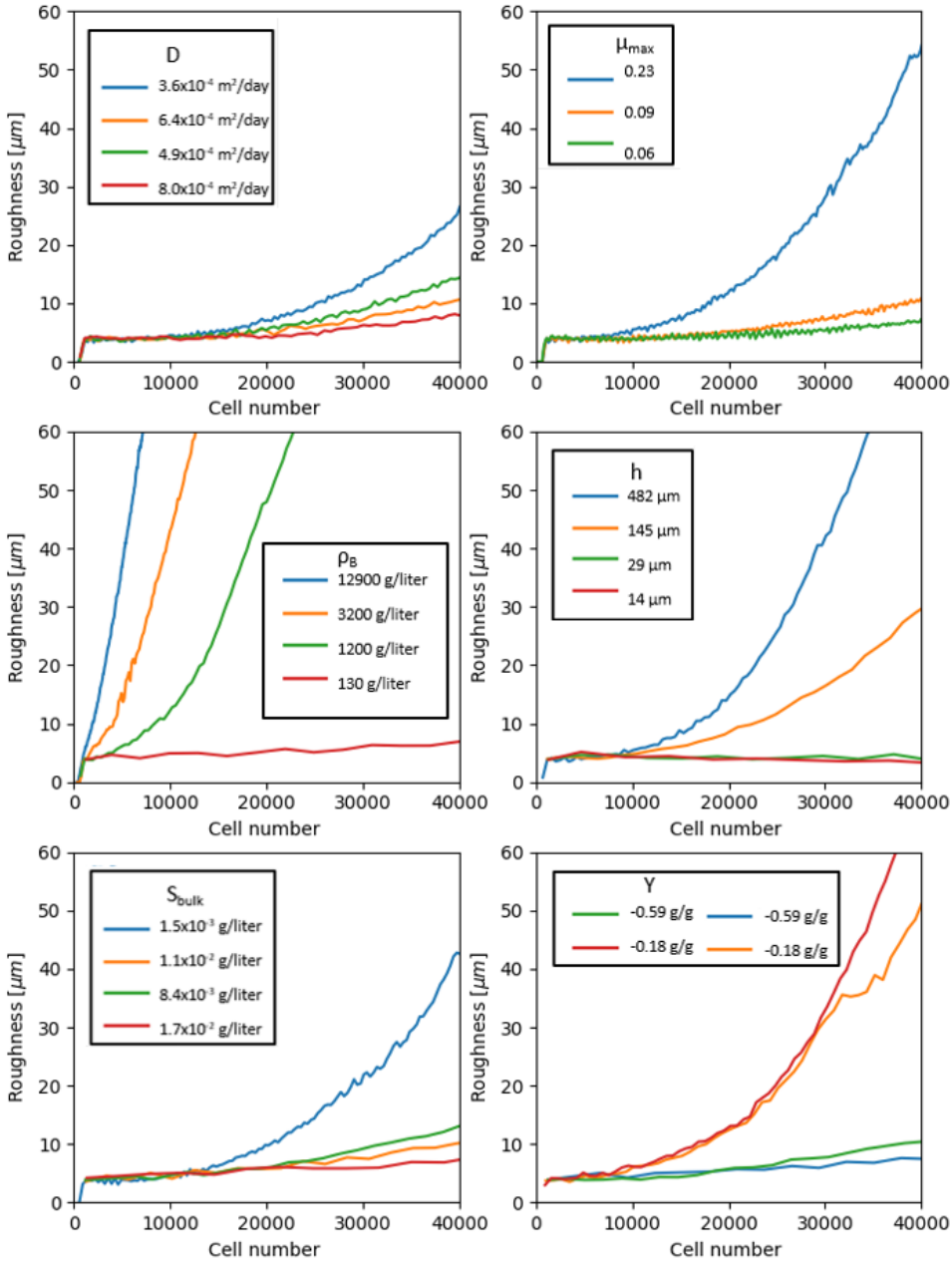



Figure 5.8: This Figure shows the trajectories of the interface roughness where the diffusivity $D$, maximum specific growth rate $\mu_{max}$, biomass density $\rho$, diffusion boundary layer height $h$, yeild $Y$ and bulk nutrient concentration $S_{bulk}$ are varied in each of the subplots respectively. The input parameters for the simulations are as in Table 3.1 unless otherwise specified in the figure legends.

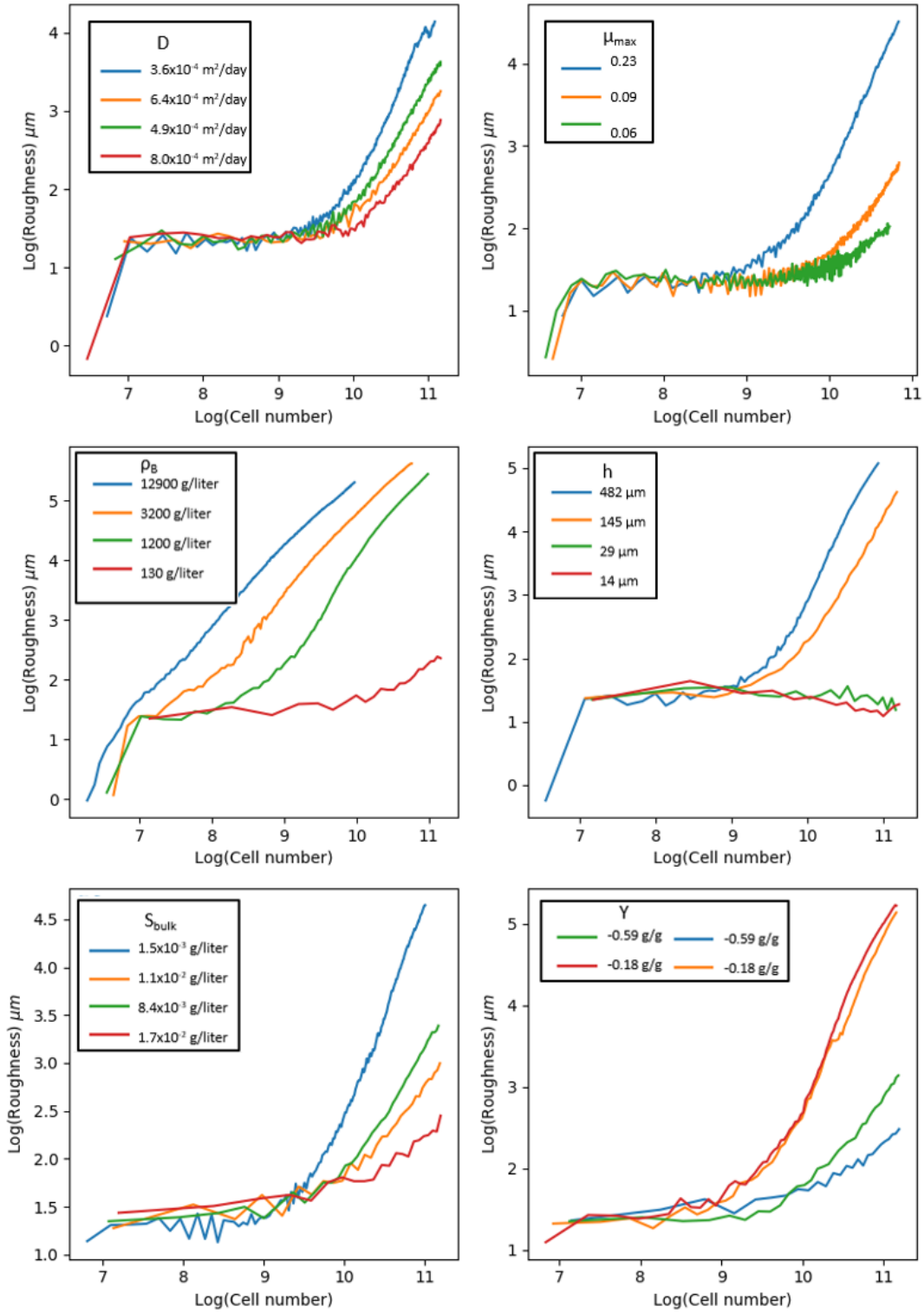



Figure 5.9: This Figure shows the trajectories of the log interface roughness against the log of time, where the diffusivity $D$, maximum specific growth rate $\mu_{max}$, biomass density $\rho$, diffusion boundary layer height $h$, yeild $Y$ and bulk nutrient concentration $S_{bulk}$ are varied in each of the subplots respectively. The input parameters for the simulations are as in Table 3.1 unless otherwise specified in the figure legends.

### 5.3.3 Long Time behaviour

As I have stated, in the rest of this chapter, I examine the long time behaviour of biofilm growth, with the aim of establishing the steady state behaviour of the interface roughness. I outlined in detail our methods for doing this in Chapter 4, and recall them briefly here for clarity. Simulations using iDynoMiCS with large numbers of cells are prohibitively slow, and so to reach long times I periodically remove cells which are not growing and are far below the growing front which are no longer influencing the growth dynamics. I do this using an algorithm which pauses the iDynoMiCS simulation after 48h of biofilm growth, clips cells below a certain height of the biofilm, and restarts the simulation again. This means I can reach long time scales without having to track an unfeasible number of cells.

Even with the use of my clipping algorithm, long time simulations remain sufficiently computationally intensive that it is necessary to carefully consider which parameter combinations to choose. It is not possible to sweep the complete parameter space as I did in Figures 5.5 and 5.8. I instead chose to vary the maximum specific growth rate $\mu_{max}$ over the range 0.1-0.4 and the bulk nutrient concentration $S_{bulk}$ over the range 0.01 to 0.0005 g/liter to produce a grid of 16 simulations. I choose these two parameters because the active layer thickness increases with $S_{bulk}$ and decreases with $\mu_{max}$, as we saw earlier in the chapter.

In order to confirm that changing these two parameters produces a range of distinct simulation outcomes, I plot a snapshot of growth rate distribution at 200000 cells of each of the simulations after a short simulation run. Figure 5.10 shows that though we are only changing two parameters $\mu_{max}$ and $S_{bulk}$, a wide variety of cell growth rate distributions are produced. There are many cells with zero or close to zero growth rate, so in the interests of showing the distribution of the most active cells I do not include these in Figure 5.10. We can see the greater the bulk nutrient concentration or the higher the maximum specific growth rate, the faster it is possible for the cells to grow. However, changing $\mu_{max}$ and $S_{bulk}$ have distinct effects on the distribution of cell growth rates within the biofilm. For high $S_{bulk}$, there can be a greater number of cells growing at high growth rates than for high $\mu_{max}$. I am therefore confident changing $\mu_{max}$ and $S_{bulk}$ produces distinct simulations and we do not effectively end up repeating simulations. that ie to test that changing the different parameters do not have the same impact on the simulations.

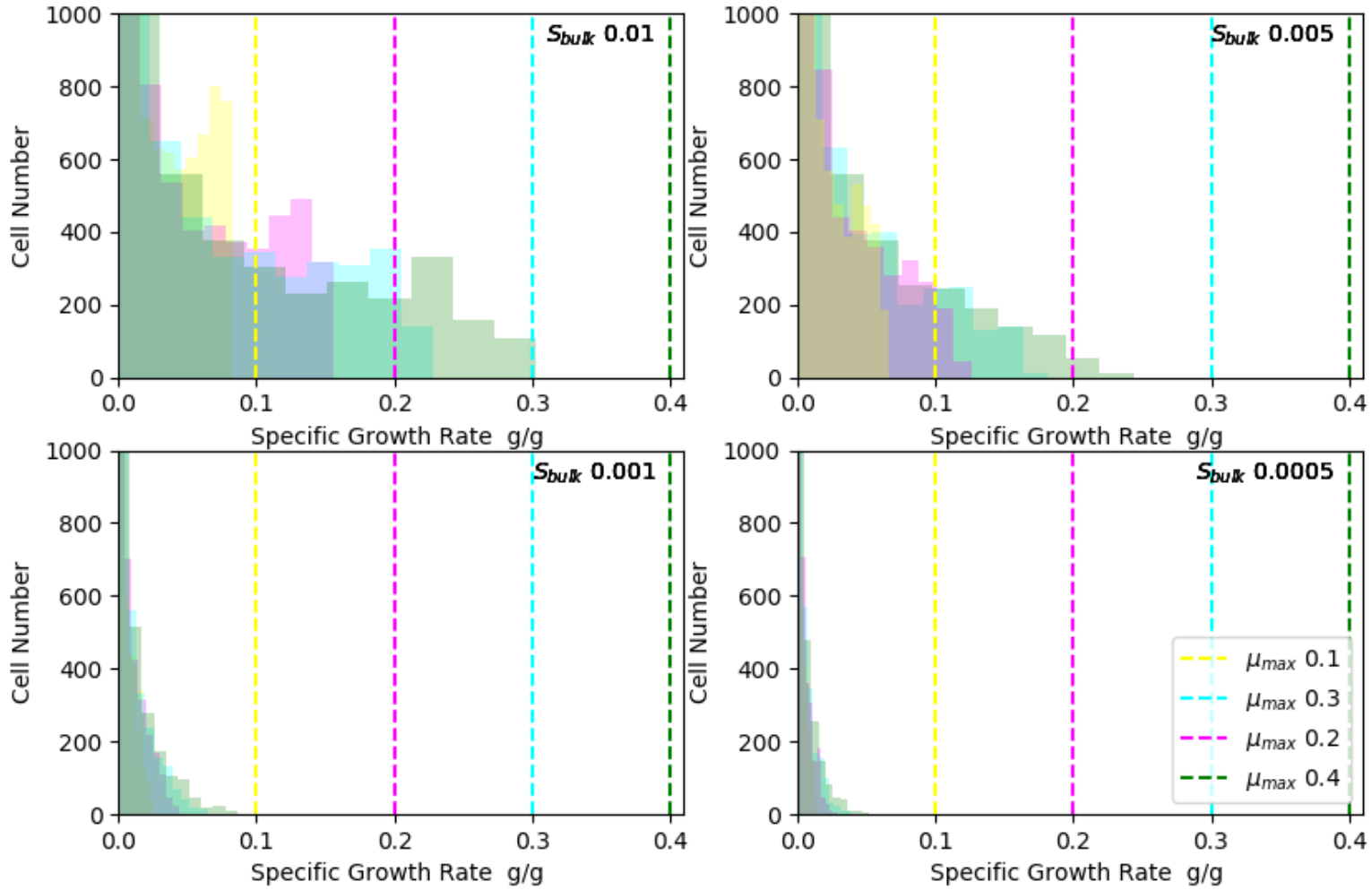



Figure 5.10: This Figure shows a histogram of the growth rate distributions of the biofilm at 200000 cells of 16 simulations with varying $\mu_{max}$ and $S_{bulk}$. Each of the subplots contains the distributions of simulations with constant $S_{bulk}$ and varying $\mu_{max}$. The y axes are adjusted such that the full number of cells with zero or close to zero growth rates are not shown on the plot.

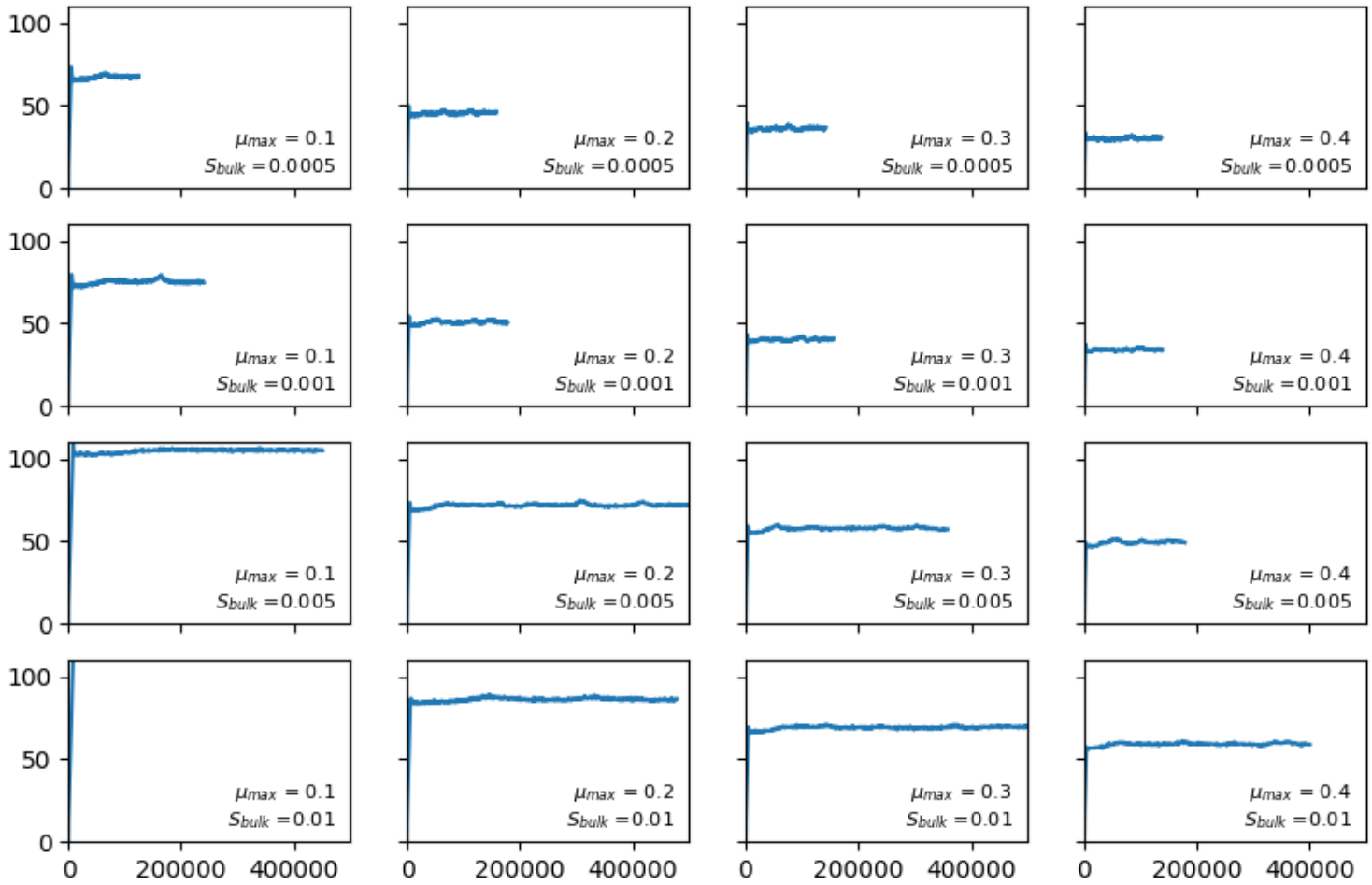



Figure 5.11: This Figure shows the active layer thickness against time for simulations with varying maximum specific growth rate $\mu_{max}$ and maximum nutrient concentration $S_{bulk}$.

I firstly confirm that the active layer thickness remains in a steady state to these long times. Figure 5.11 shows the average active layer thickness plotted against time for simulations with varying $S_{bulk}$ and $\mu_{max}$. We can see the average active layer thickness reaches a steady state as the biofilm has formed and remains constant to long times. We can see that increasing the bulk nutrient concentration $S_{bulk}$ increases the thickness of the active layer, because there are more nutrients available. We can also see that increasing the specific cell growth rate decreases the active layer thickness, as for a constant yield increasing growth rate means more nutrient consumption per cell, so less cells are able to grow leading to a thinner active layer.

I next consider the trajectories of the interface roughness in Figure 5.12. We can see that some of the roughness trajectories now reach a steady state, though others continue to increase at these long times. On closer inspection, the trajectories of the interface roughness can broadly be classified into three types. The first, for large $S_{bulk}$ and small $\mu_{max}$ we can see that the interface roughness reaches a steady state at a low roughness. There are also interface roughness trajectories which continue to increase for long times for low $S_{bulk}$ and large $\mu_{max}$. There also appear to be roughness trajectories with a fluctuating steady state behaviour at some intermediate values. Comparing these trajectories to the short time simulations in Figure 5.8, we can see that the steady states we saw at this time scale are likely not the true steady states, as the roughness continues to increase for all the simulations we see here.

The distinctiveness of these three classes of trajectories is more apparent when they are plotted together on the same plot in Figure 5.13(a). We can see that the roughness trajectories which reach a constant steady state correspond to rather smooth biofilms (low average roughness), the trajectories that are monotonically increasing correspond to biofilms with high roughness, while the fluctuating phase is an intermediate phase. It is also then apparent that the trajectories that do not fit into any of these three classes may be transitioning between two phases. We discuss this transitional behaviour in more detail in the next chapter, and focus for the rest of this chapter on the remaining trajectories in Figure 5.13(b).

Finally, I also investigate the behaviour of the the standard deviation of the active layer thickness. We saw in Section 5.2 that the active layer thickness was calculated for vertical strips of width 8 $\mu m$ across the biofilm. The standard deviation of the thickness of each of these strips is then taken to calculate the standard deviation of the active layer thickness. Figure 5.14 shows us the standard

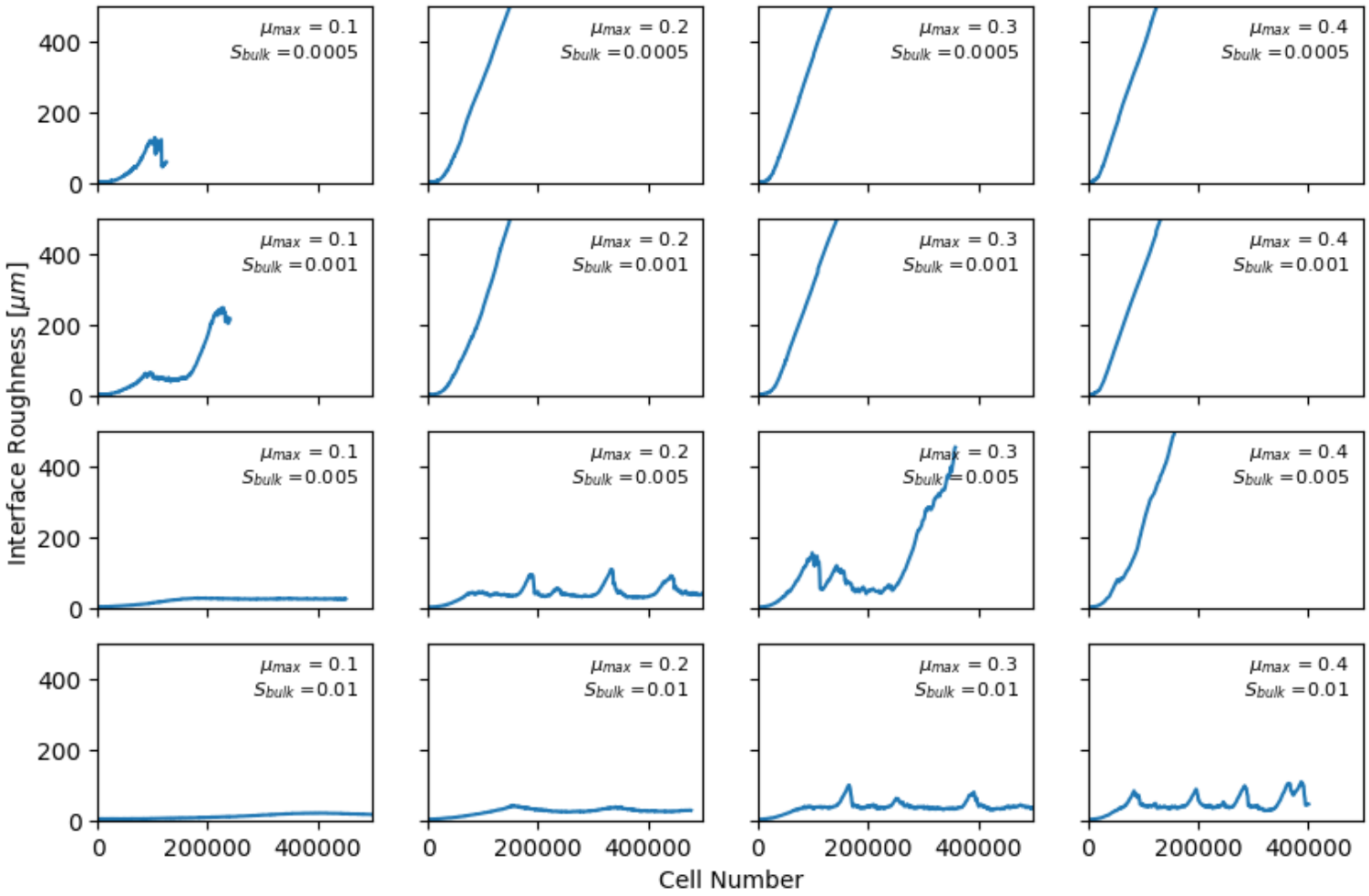
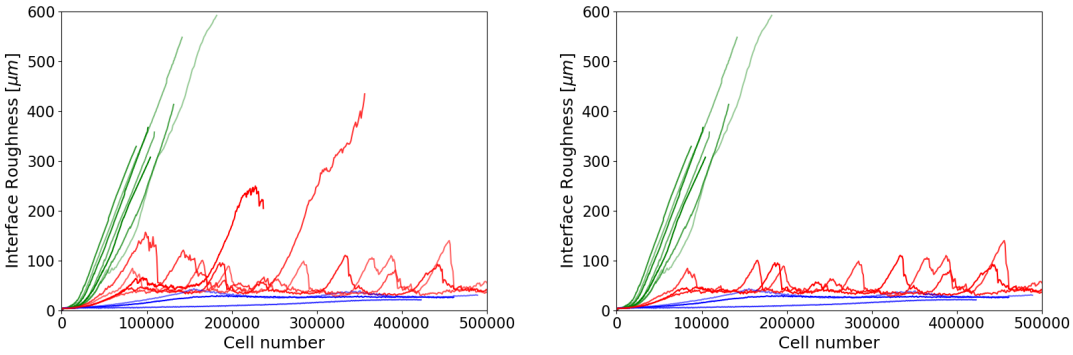



Figure 5.12: This Figure shows the interface roughness against time for simulations with varying maximum specific growth rate $\mu_{max}$ and maximum nutrient concentration $S_{bulk}$. The other input parameters for the simulations are as in Table 3.1.




(a) All Roughness (b) Transitions excluded

Figure 5.13: This Figure shows the interface roughness for the set of simulations with varying bulk nutrient concentration and maximum specific growth rate that I have shown in Figures 5.11,5.12 and 5.14. The left hand panel shows the trajectories for all 16 of the simulations shown previously in Figure 5.12, and the right hand panel excludes the transitional trajectories with $\mu_{max} = 0.3$, $S_{bulk} = 0.005$ and $\mu_{max} = 0.1$, $S_{bulk} = 0.001$. The trajectories with a monotonically increasing roughness are coloured green, the trajectories with a fluctuating steady state are coloured red and the trajectories with a smooth steady state are coloured blue.

deviation of the active layer thickness for each simulation. We can see that the variation in the thickness of the active layer can be significant compared to the average active layer thickness; for example, the simulation with $\mu_{max} = 0.2$, $S_{bulk} = 0.001$ can have an active layer thickness standard deviation as much as $25\mu m$, compared to a average active layer thickness of $75\ \mu m$. It is also apparent that though the standard deviation of the active layer thickness grows initially, it reaches a steady state, albeit one which fluctuates significantly in many cases. The time (shown here in terms of cell number) at which the steady state of the active layer thickness standard deviation is reached is also significantly later than average active layer thickness steady state and is similar to the time the roughness reaches a steady state. This is something we will investigate in detail in the later chapters of this thesis.

### Time v Cell number

I now return to the question of why I plotted the quantities against the cell number rather than time. This means I chose to compare biofilms of the same size rather than biofilms that have been growing for the same amount of time.

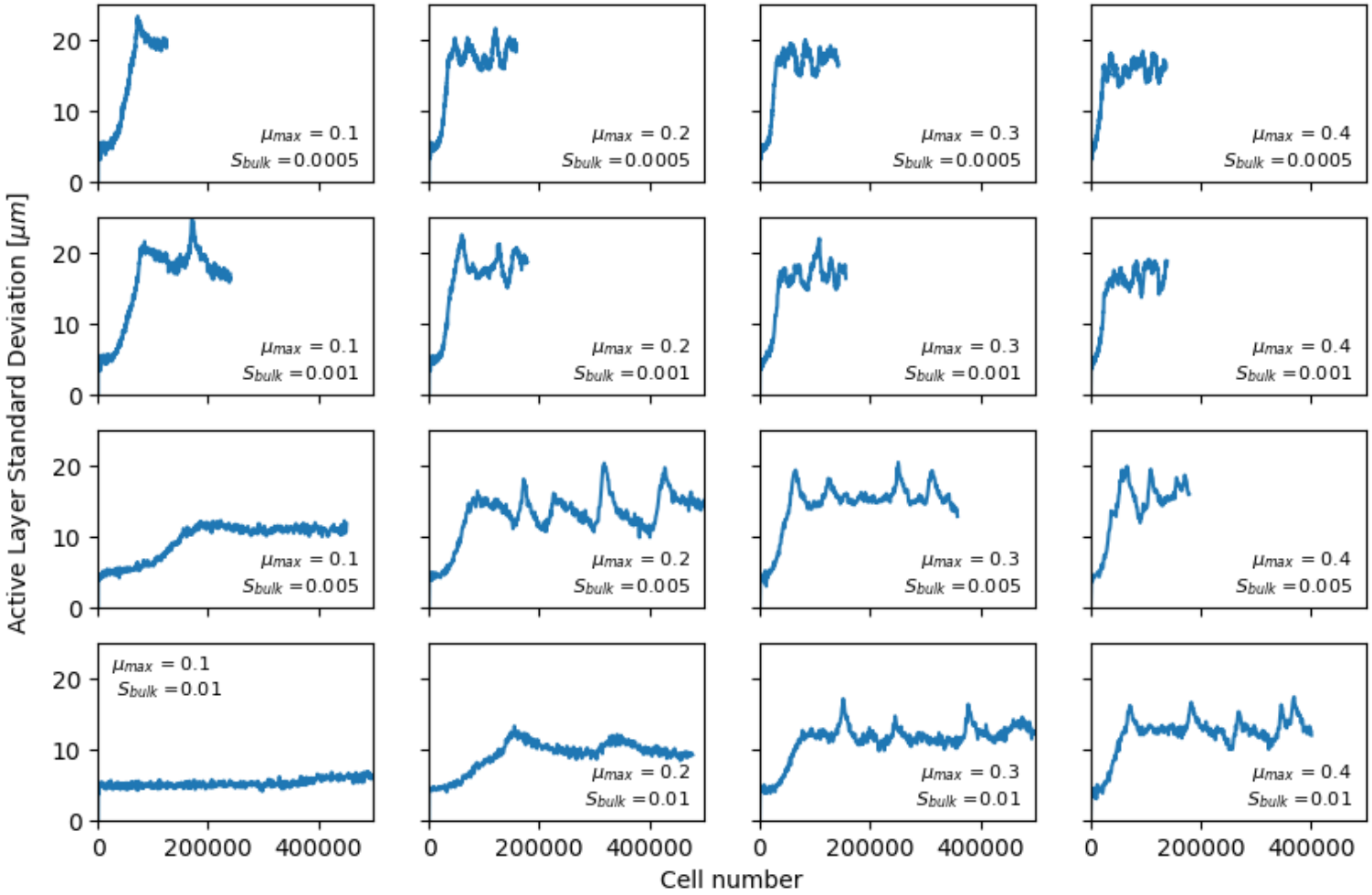
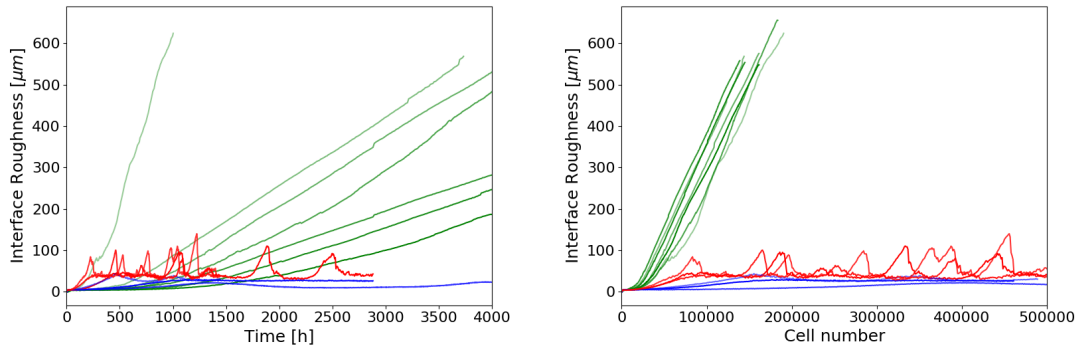



Figure 5.14: This Figure shows the standard deviation of the active layer thickness against cell number for simulations with varying maximum specific growth rate $\mu_{max}$ and maximum nutrient concentration $S_{bulk}$. The other input parameters for the simulations are as in Table 3.1.




(a) (b)

Figure 5.15: This Figure shows the interface roughness to long simulation times. The same simulations are shown in both sub-figures, the only difference is that (a) is plotted against time and (b) is plotted against cell number. The trajectories with a monotonically increasing roughness are coloured green, the trajectories with a fluctuating steady state are coloured red and the trajectories with a smooth steady state are coloured blue.

In Figure 5.15, we can see the long time roughness trajectories plotted together with both time and cell number on the x-axis. It is clear that the different types of trajectories can be more clearly distinguished for different biofilm sizes rather than growth times. We can see that for biofilms with the monotonically increasing roughness (or roughest phase), can still have a roughness that is less than the smooth steady state (or smooth phase) up to to 1500 hours of growth. Though plotting the interface roughness against time is more conventional, authors such as Drescher *et. al.* also find plotting against cell number to be useful [102].

## 5.4 Discussion

In this chapter I investigated the trajectories and steady state behaviour of the active layer thickness and the interface roughness. We saw that the average active layer thickness reaches a constant steady state at very early times. We also saw that reaching the steady state of the interface roughness requires the use of a long-time simulation algorithm. We observed three distinct behaviours - a low roughness steady state, an intermediate roughness fluctuating steady state and a monotonically increasing relaxation behaviour. I focus my discussion here on the implications this work has for characterising the phase behaviour of biofilm growth.

As we have seen, the active layer thickness reaches a constant steady state for all parameter sets I have surveyed. This observation is consistent with several other authors who have investigated the active layer, both in experiments and simulations [28, 42, 64, 65]. We saw in Chapter 2 that Nadell *et. al.* predict the steady state active layer thickness to be correlated with a combination of the simulation parameters $\delta = \sqrt{S_{bulk} D_G Y / \mu_{max} \rho h^2}$ [28]. In my work, we saw that these key parameters do indeed influence the active layer thickness, but that they do not have the relationship with the active layer thickness that Nadell *et. al.* expect. Though the key input parameters in the iDynoMiCS simulation software are the same as the parameters which make up Nadell's $\delta$ parameter, Nadell *et. al.* are not using the same simulation software as I am. Additionally, in my work I established the key input parameters of iDynoMiCS by examining the input file, rather than deriving an expected active layer parameter from the iDynoMiCS reaction-diffusion equations. This would be non-trivial due to the hybrid nature of the individual-based simulations, but I think a worthwhile endeavour given the importance of the active layer in bioiflm spatial structure which I will continue to discuss throughout this thesis.

We also saw that the interface roughness was hard to characterise at early times in our simulations. I argue that this is comparable with the challenges in characterising the early time roughness seen in interface growth theory studies of biofilm and colony growth. While interface growth theory in general characterises interface growth behaviour using exponents for the early ($\beta$) and late ($\alpha$) time roughness behaviour, for biofilms it is usual to only use the late time $\alpha$ exponent to characterise growth. It is commonly argued that comparing early time roughness exponents of classic interface growth theory to biofilms is not appropriate because of the time it takes the biofilm to form - i.e. the interface is not beginning from a continuous uniform interface as in classical interface growth theory. I argue that the same is true in my analysis. While I did not measure the growth exponents of the interface roughness in my analysis, we did see that in the early stages of biofilm growth the roughness trajectories are not linear on a log-log plot, and so motivate our long-time simulations where the relaxed behaviour could be seen.

Once I had produced simulations that reached long simulation times, we saw there was not a clear steady state in all the phases, and there are instead three distinct long-time behaviours. While these simulations took significant work to produce, I think that their importance in allowing me to identify distinct phases and allowing me to compare them to the rest of the literature justifies this. In

Chapter 2, we saw that the long-time behaviour of the interface roughness was important in classifying the phase behaviour - either so the scaling exponents could be calculated from the steady state or so the monotonically increasing roughness could indicate the transition to a possible diffusion-limited aggregation (DLA) phase. We also saw in Chapter 2 that the steady state roughness behaviour is also used in the reaction-diffusion literature, though the long time behaviour is not always reached. I believe this means that the interface roughness is not always characterised in a consistent manner as it is not clear what stage of growth it is in. For example, Farrell *et. al.* observe a steady state roughness at very early growth times [16]. This could be due to the different types of simulations that are used there, but it is also possible they are measuring the pseudo-steady state I see at early times in my simulations before the true steady state is reached.

While I have argued that long time simulations are necessary for characterising the biofilm spatial structure, it is also necessary to discuss the biological implications of such long time scale simulations. Although I am simulating biofilm growth in a flow-cell like set up, biofilms which are 500,000 cells large create very tall thin biofilms of the kinds of sizes which are not seen in flow cells. However, a procedure akin to our long time simulation algorithm (clipping) is not unprecedented experimentally. For example, in oral biofilm research moving the top portion of a growing biofilm from one reactor to another is an established experimental procedure [103]. Outside of specific flow cell set up, we saw in Chapter 2 that large population range expansions are a common scenario with significant evolutionary implications. While the uninterrupted growth of such a large biofilm can not be said to be a likely *in vivo* or *in vitro* scenario, studying the growth in this way is a useful neutral case for what would have happened during a rage expansion without environmental interruptions.

We also saw that in contrast to the interface fluctuations (roughness), the active layer thickness fluctuations (active layer thickness standard deviation) reached a steady state in all the long time simulations produced. Since the active layer is along the moving interface, a decoupling of the active layer thickness fluctuations and the interface fluctuations suggests that gaps in the active layer may be a cause of this. Despite these differences, other aspects of the trajectories of the roughness and the active layer thickness standard deviation are remarkably similar. Firstly, steady states which are smooth for the roughness are also smooth for the active layer thickness standard deviation, and the same goes for the fluctuating steady states. Secondly, the times at which those steady states appear is also similar.

Taken together, this implies a close relationship between the interface roughness and the active layer thickness. In the next chapter, I examine this relationship in detail and are able to elucidate the behaviour of the roughness.

## 5.5 Summary

In summary, in this chapter I outlined my methods for calculating the active layer thickness and the interface roughness and some initial results using these calculation methods. I find that the average active layer thickness reaches a steady state as soon as the biofilm has formed, and this average active layer thickness has unique relationship with each of the key simulation input parameters. I also found that the interface roughness can not be easily characterised at short time scales, and discuss in detail my motivations for performing long time simulations. I find that at long time scales, the interface roughness has three distinct interface roughness behaviours, a smooth steady state, a fluctuating steady state and a monotonically increasing long-time state.

# Chapter 6

# Interface Pinning and Biofilm Roughness

## 6.1 Outline

In the last chapter, we saw that to understand the phase behaviour of a growing biofilm it is necessary to study its long time behaviour. We also saw that the interface roughness at long times had three distinct behaviours depending on the parameters of the simulations - a smooth steady state, a fluctuating steady state and a monotonically increasing behaviour without a steady state. We also saw that although the average active layer thickness reaches a constant value after the biofilm has formed, the active layer thickness can vary significantly across the biofilm width. In this chapter, I examine the behaviour of the active layer in detail and relate it to the behaviour of the interface roughness.

I begin this chapter by outlining some additional analysis methods: in particular the definition of the stationary interface fraction and the inactive interface fraction. I then describe how my simulations show interface pinning behaviour, in which parts of the interface become stationary and lag behind the growing interface, sometimes on a permanent basis and sometimes temporarily. Examining this phenomenon in detail allows me to understand how the different categories of interface roughness behaviour that discussed in Chapter 5 are produced. I finish the chapter by discussing how this pinning behaviour might relate to the pinning-depining transition known in interface growth theory, as well as how the

active layer behaviour relates to studies of the active layer in the reaction-diffusion modelling literature.

## 6.2 Analysis Methods

### 6.2.1 Inactive Interface Fraction

This chapter uses many of the same analysis methods outlined in the previous chapter, though we also introduce several new analysis methods in this chapter. The first of these is the inactive interface fraction. We recall from Chapter 5.3 that to calculate the position of the interface I use a grid with the average biomass, growth rate etc of the cells in each grid square. I define the interface by finding which of these grid squares have non-zero biomass and are adjacent to grid squares which are empty of biomass. These are the coloured squares in Figure 6.1. In this chapter, I further define the inactive interface. These are the blue squares in Figure 6.1(a) which are the grid squares that have an average specific growth rate which is less than the active layer cutoff $\mu_{AL}$,

$$\mu_{AL} < \frac{\mu_{max}}{1000} \frac{S_{bulk}}{k_s + S_{bulk}} \tag{6.1}$$

where $\mu_{max}$ is the maximum specific growth rate, $k_S$ is the solute concentration at which the growth is half maximal and $S_{bulk}$ is the bulk solute concentration (see Equation 5.3 in Chapter 5).

I then calculate the inactive interface fraction $f_I$ for a particular configuration by dividing the number of inactive interface grid squares $N_I$ by the total number of interface grid squares $N_{total}$. Simply,

$$f_I = \frac{N_I}{N_{total}}. \tag{6.2}$$

### 6.2.2 Pinned Interface Fraction

The second new definition I introduce in this chapter is the stationary interface fraction, which can also be called the pinned interface fraction. To find the parts of the interface which are pinned, I compare the position of the biofilm interface six hours apart. The calculation involves comparing the positions of the interface

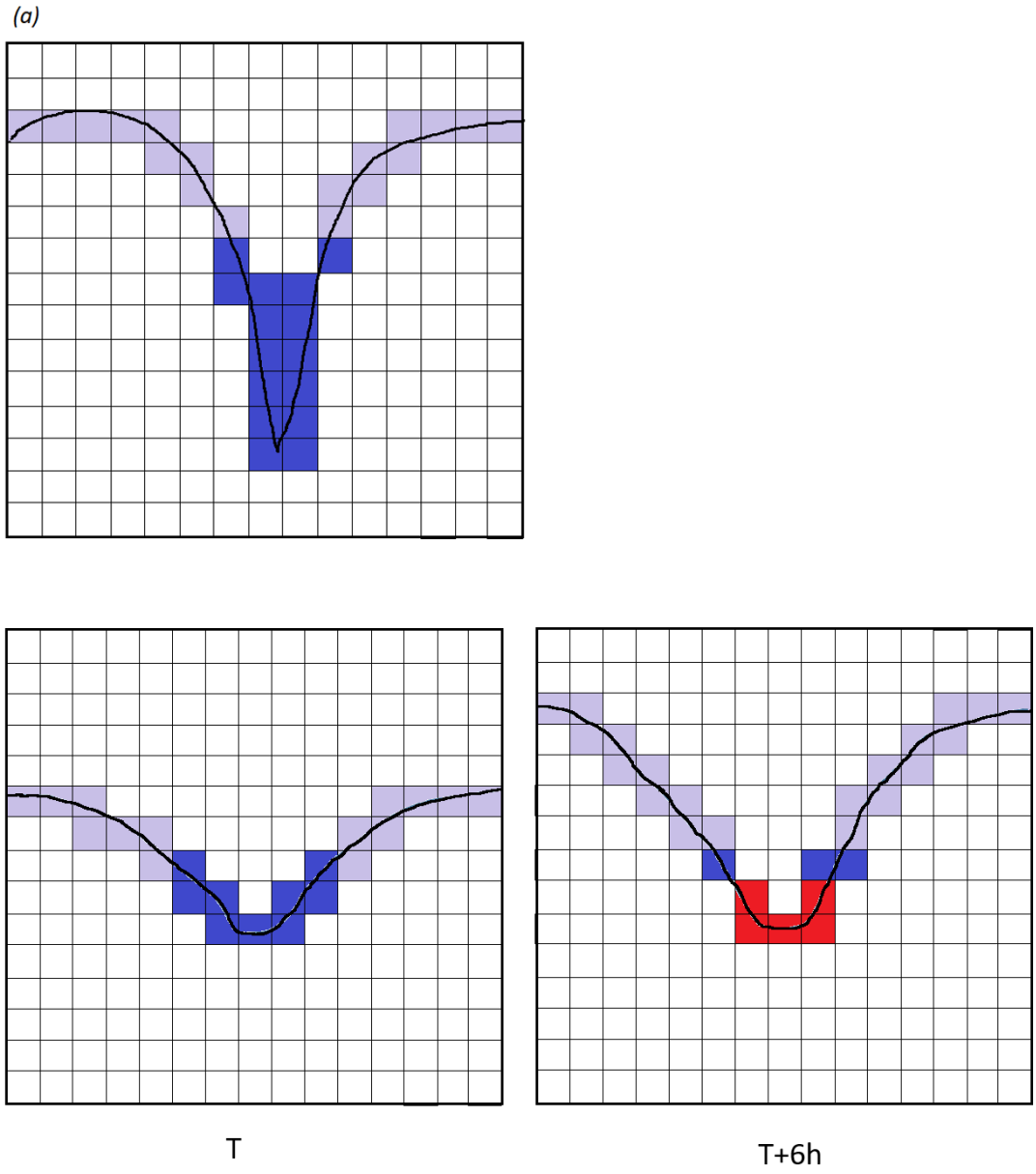



Figure 6.1: Sketches of the inactive and pinned interface points. All coloured squares are part of the interface. Sketch (a) shows the inactive interface coloured in blue - these are the parts of the interface where the average growth rate is less than the condition for being in the active layer, as discussed in the text. Sketch (b) shows two interface sketches 6 hours apart. Again, the interface squares which are inactive are shown in blue. Comparing the grid squares that comprise the interface in the two panels, we can see that some of them remain the same. These are the pinned interface grid squares, shown in red.

grid squares, and finding the interface grid squares which are common to both configurations. These are the red squares in Figure 6.1(b), which we can see are common to the configurations in the left and right panels which are 6 hours apart. I chose to compare the interface six hours apart as this is the frequency of output files and so the minimum time period I could have chosen. The minimum time period is most appropriate because once parts of the interface have become stationary, they do not move again, so this allows us to most accurately capture changes in the pinning interface fraction. An additional subtlety in the calculation is due to the resolution of the iDynoMiCS grid squares ($8\mu m$). Some very slow growing interface points do not move a whole grid square in 6h, meaning there there are some interface points which are common to configurations 6h apart but are part of the growing front. To remove these spurious points from our calculation and find the interface points which are truly pinned, I do only count interface squares which are inactive as part of the pinned interface, as defined in the previous section.

I then calculate the fraction of interface points which are both inactive and have been stationary for the previous 6h. As such, the pinned interface fraction $f_P$ is defined as,

$$f_P = \frac{N_P}{N_{total}} \tag{6.3}$$

where $N_P$ is the number of inactive interface points which have not moved in 6 hours.

## 6.3 Results

### 6.3.1 Active Layer Dynamics and Interface Pinning

I begin by looking more carefully at the active layer dynamics. We saw in the last chapter that the average active layer thickness reaches a constant steady state soon after the biofilm is formed, while the standard deviation of the active layer thickness only reaches a steady state much later on. This implies that local variations in active layer thickness along the interface take much longer to reach steady state than the average value. Figure 6.2 shows an example of the trajectory of the average active layer alongside snapshots of the biofilm's growth. It is clear that there can be many different biofilm configurations with large local variation in the active layer thickness, for the same average active layer thickness.

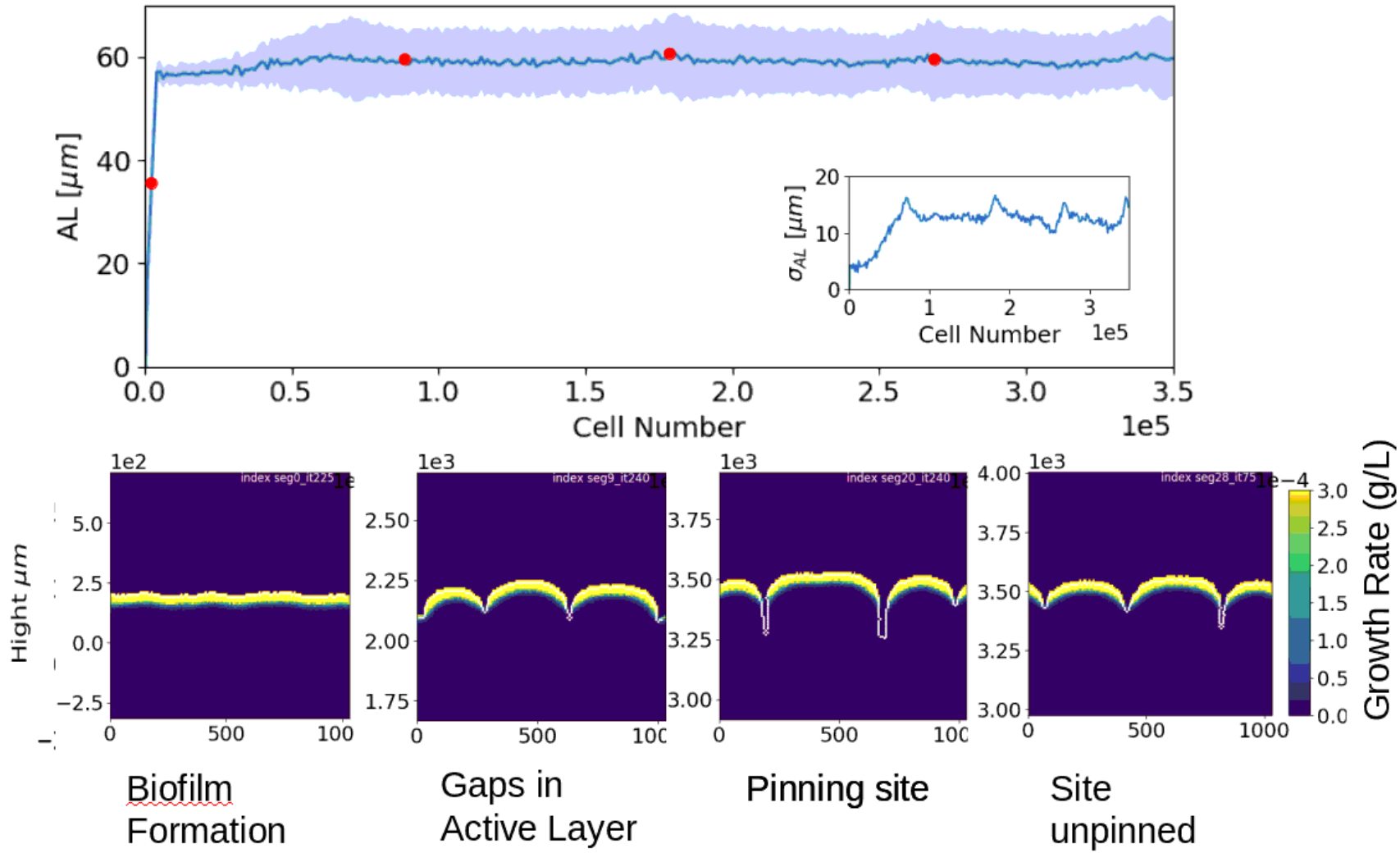



Figure 6.2: This Figure shows the trajectory of the active layer thickness and the standard deviation of the active layer thickness for an example biofilm simulation. These plots are shown alongside snapshots of the biofilm configuration at the time points shown as red dots on the active layer thickness plot.

It is this dynamical behaviour of the local active layer thickness which I focus on in this section, rather than the average active layer thickness.

Returning to the grid of simulations for varying $\mu_{max}$ and $S_{bulk}$ which we looked at in the last chapter, I now plot the active layer thickness across the width of the biofilm in Figure 6.3. For each subplot, the cell number is on the x-axis and biofilm width is on the y-axis. We recall from Chapter 5 that the active layer thickness is calculated for $8\mu m$ bins across the width of the biofilm, and thus the color of each subfigure represents the local active layer thickness at that point in time and space. We can see from the scale that darker colours represent thin or no active layer and the lighter colours representing a thicker active layer. This means that as we move from that moving from left to right on each subplot, the pattern of dark lines we can see represent the progression of active layer gaps, or where parts of the active layer is thin, over the course of biofilm growth. As an example, we see in Figure 6.4 an example of the active layer colourmap ($\mu_{max} = 0.3$, $S_{bulk} = 0.01$) alongside snapshots of the biofilm's growth to demonstrate that the dark lines in Figure 6.3 really do represent thin parts of the active layer or gaps moving around.

Looking carefully at these different patterns in Figure 6.3, we can see there is significant diversity in the behaviour of the active layer for different simulation parameters. With careful examination we can see that there are broadly three categories of behaviour. Example trajectories representing each category are plotted in Figure 6.5. Firstly, there is the 'smooth phase' in which the active layer remains thick and fairly uniform across the biofilm width. Secondly, there is a phase where we can observe thin portions of the active layer which move around, and produce a distinctive pattern. Finally, there is a phase where there appear to be active layer gaps which are wide and appear to engulf parts of the active layer which have become locally thin.

I now look at snapshots of the biofilm configuration for each of these three phases. In Figure 6.6, we can see snapshots of the configurations, taken when they have reaches sizes of approximately 25000, 50000, 75000 and 100000 cells. On the top panel, we can see the progression of a biofilm that retains a smooth interface throughout and has a thick active layer. In the bottom panel, we can see the biofilm configuration which is very rough and develops large, distinct fingers which remain throughout the trajectory of the biofilm. The troughs of these fingers that remain stationary I refer to as pinning sites. In the central panel, we can see biofilm trajectory which develops gaps during its growth, and parts

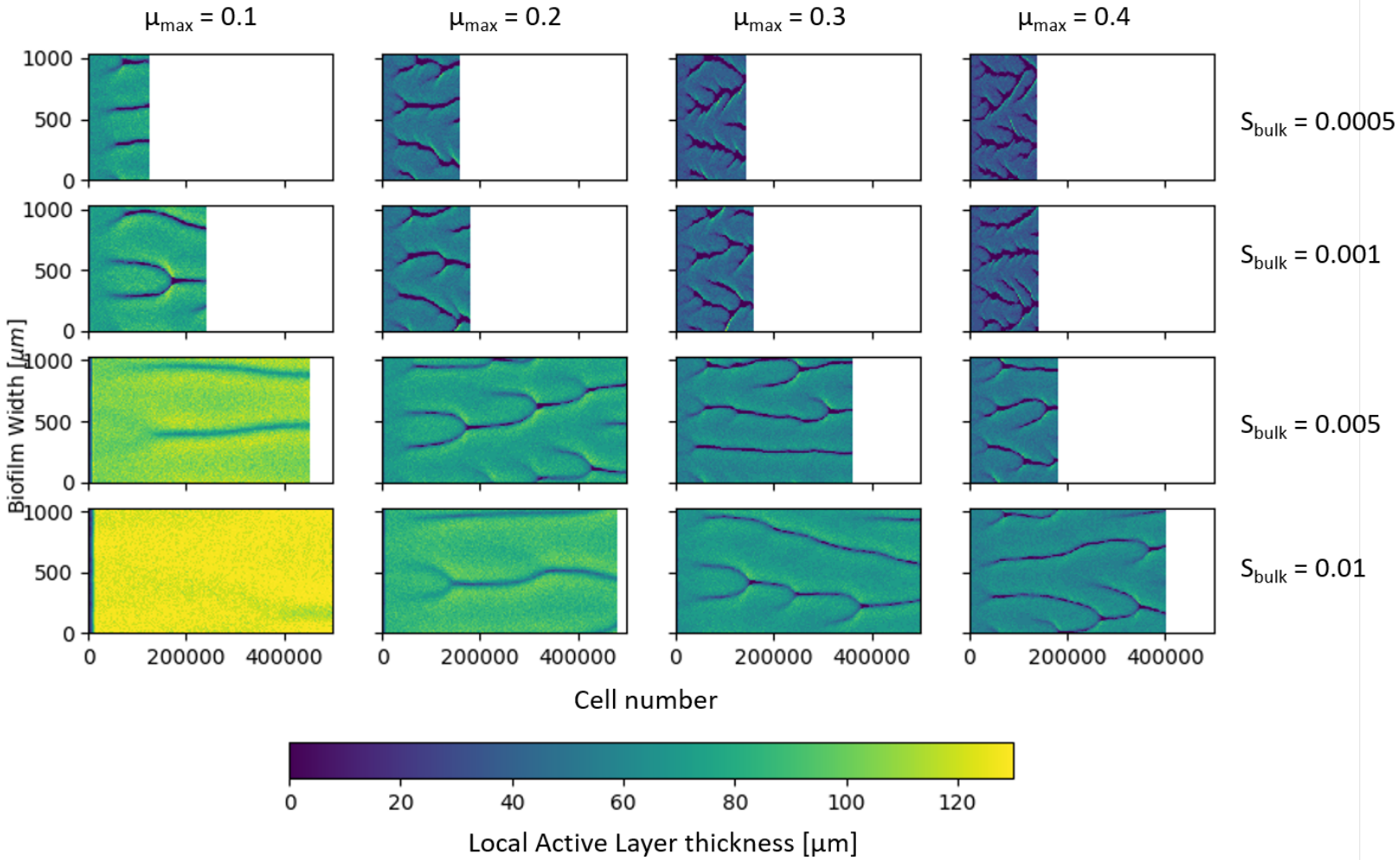



Figure 6.3: This Figure shows colourmpaps of the local active layer thickness for simulations with varying maximum specific growth rate $\mu_{max}$ and maximum nutrient concentration $S_{bulk}$. The other input parameters for the simulations are as in Table 3.1. For each subplot, the x-axis is cell number, the y-axis is the position along the biofilm width and the colour of the plot is the local active layer thickness.

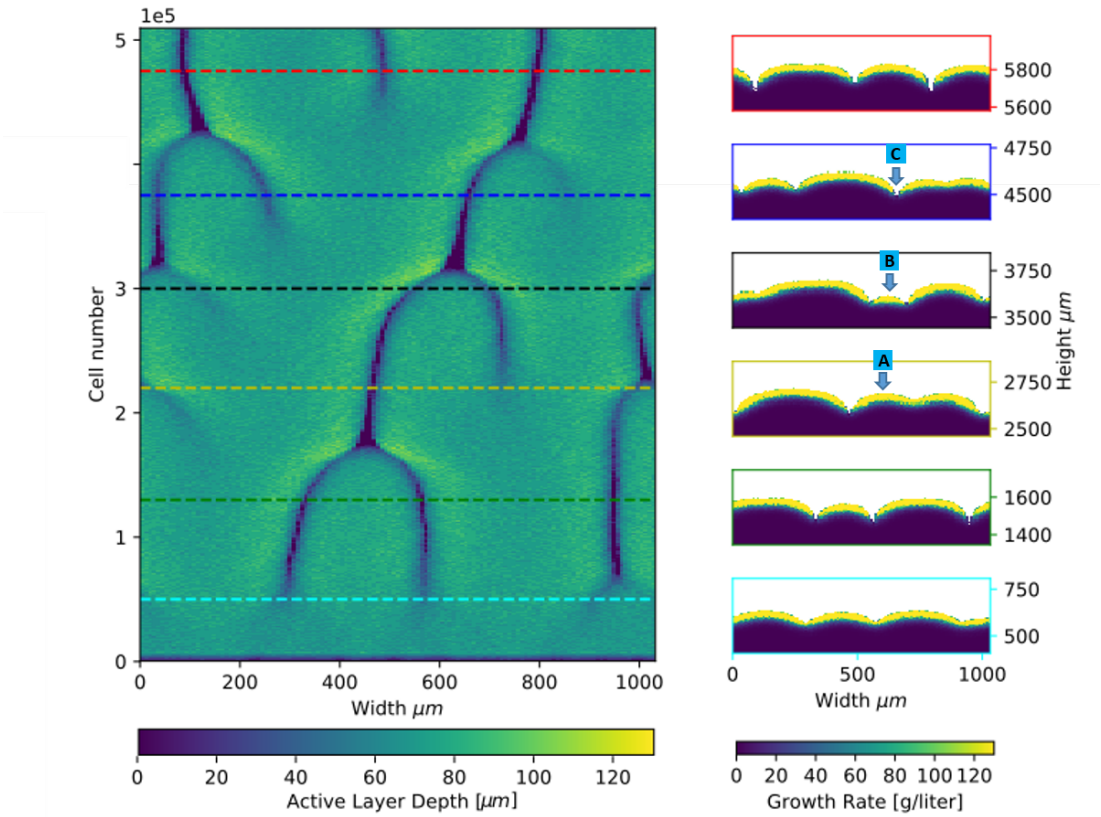



Figure 6.4: On the left hand side of this figure is a colourmap of the local active layer thickness across the width of the biofilm over the course of biofilm growth. On the right hand side of the figure is snapshots the biofilm configurations. The colour of the dashed lines on the left hand plot each have a corresponding biofilm snapshot for the same cell number on the right hand side with a border of the same colour.

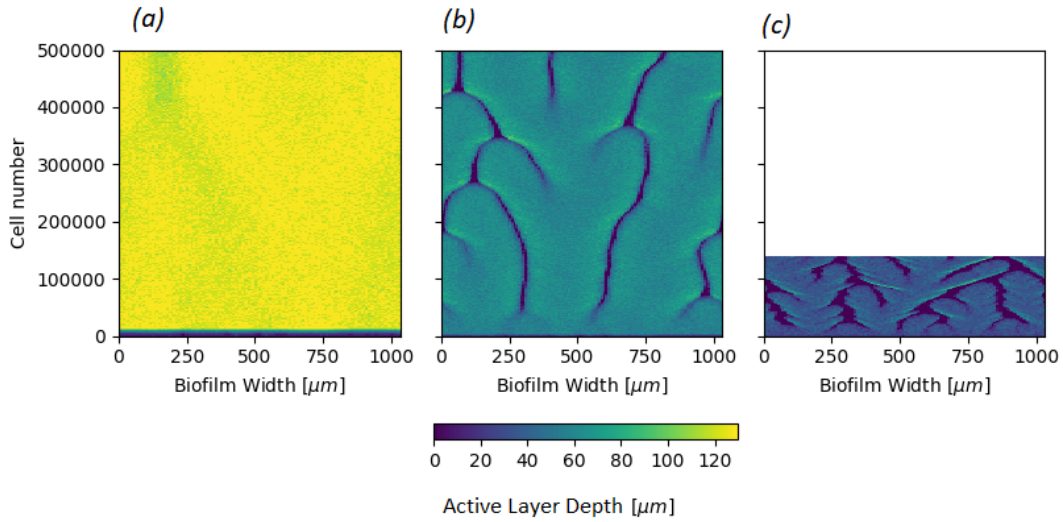



Figure 6.5: This Figure shows three of the active layer thickness colourmaps from Figure 6.3 which represent three qualitatively distinct types of behaviour as discussed in the text. Figure (a) is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.1$, Figure (b) is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.4$ and Figure (c) is for parameters $S_{bulk} = 0.0005$, $\mu_{max} = 0.4$.

of the interface which lag behind the interface for periods of time. These troughs which remain stationary for a finite period of time I refer to as depinning sites, because they pin and then depin over the course of the biofilm trajectory.

The snapshots in Figure 6.6 show the three 'phases' can be distinguished not just by the roughness of the biofilm interface but also whether or not the active layer has gaps, and whether or not parts of the active layer are stationary. For the rest of this thesis, I refer to each of these three phases as the unpinned, depinned and pinned phases respectively, by analogy with the interface growth theory literature we discussed in Chapter 2, where a pinning-depinning transition is well known. Specifically, the pinned phase is when pinning sites occur and remain, depinning is when pinning sites appear but can be overcome, while in the unpinned (or 'flat') phase, there are no pinning sites. I will discuss whether there is a quantitative relationship between my phases and those known in interface growth theory literature in the discussion of this chapter.

The distinction between the three phases is further emphasized if we look at the parts of the interface which are stationary. Figure 6.7 shows the same snapshots as Figure 6.6 but this time the inactive interface points are coloured yellow, while the stationary parts of the interface are coloured red, as defined in Section 6.2.

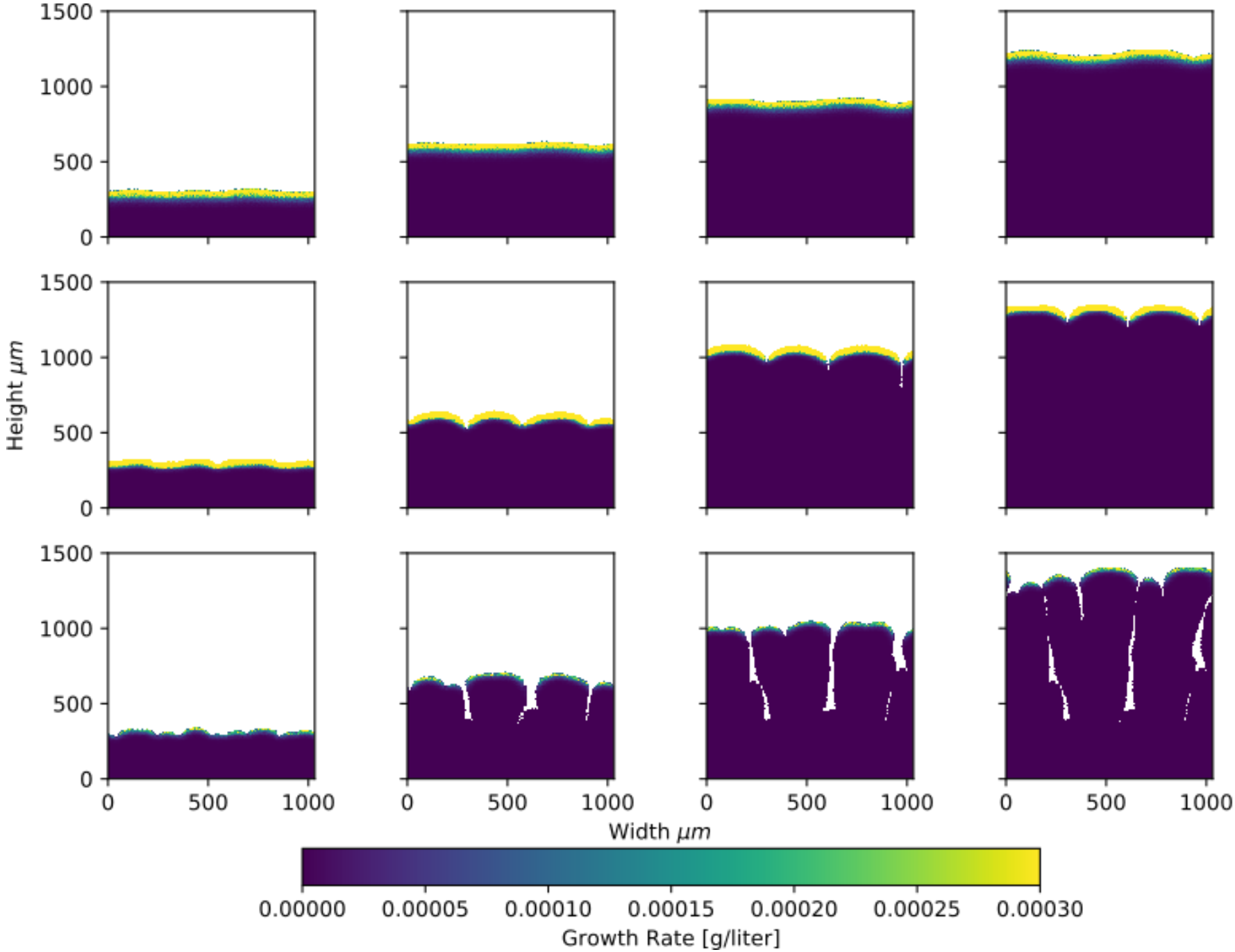



Figure 6.6: This Figure shows snapshots of the biofilm configurations for simulations that represent each of the three phases. The top row of snapshots is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.1$ which represent the 'smooth' phase described in the text. The central row of snapshots is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.4$ which we refer to as the depinned phase in the text. The bottom row of snapshots is for parameters $S_{bulk} = 0.0005$, $\mu_{max} = 0.4$ and represents the pinned phase. Each column of snapshots is for

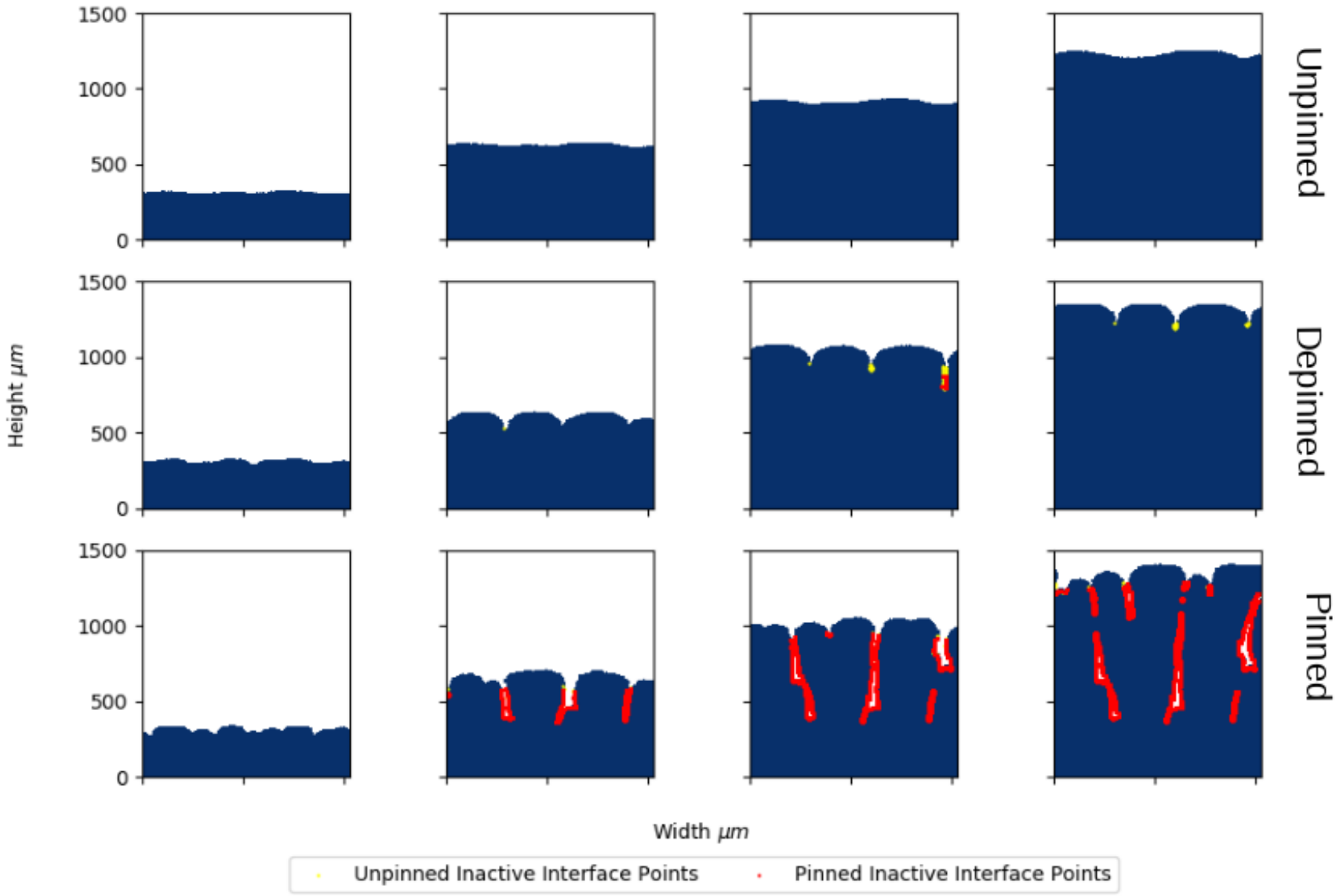



Figure 6.7: This Figure shows snapshots of the biofilm configurations for simulations that represent each of the three phases. The yellow dots are parts of the interface which are inactive. The red dots are parts of the interface that are inactive and stationary (or in other words pinned as I define in the text). The top row of snapshots is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.1$ which represent the 'smooth' phase described in the text. The central row of snapshots is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.4$ which we refer to as the depinned phase in the text. The bottom row of snapshots is for parameters $S_{bulk} = 0.0005$, $\mu_{max} = 0.4$ and represents the pinned phase. Each column of snapshots is for biofilms with 25000, 50000, 75000, 100000 cells.

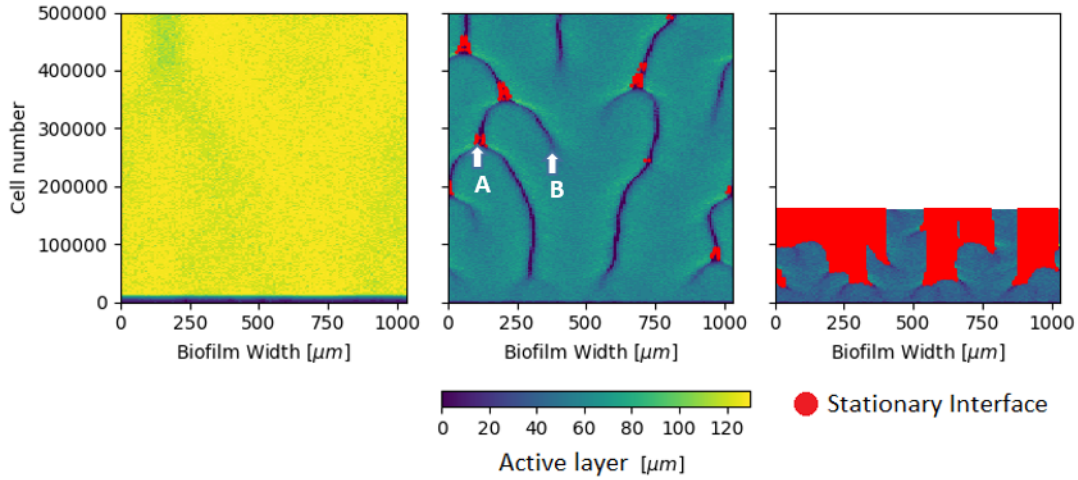



Figure 6.8: This Figure shows three of the active layer thickness colourmaps from Figure 6.3 which represent three qualitatively distinct types of behaviour as discussed in the text. Figure (a) is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.1$, Figure (b) is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.4$ and Figure (c) is for parameters $S_{bulk} = 0.0005$, $\mu_{max} = 0.4$. The red dots represent part of the interface which is stationary, or in other words pinned, as defined in Section 6.2. The significance of labels A and B will be discussed in Section 6.3.3.

My definitions of the inactive and stationary interface are outlined in Section 6.2. Figure 6.8 shows the the stationary interface points plotted on top of the active layer colourmaps for the same simulations. As we expect, there are no stationary interface points in the unpinned phase. In the depinned phase, interestingly the stationary interface points (and therefore pinning sites) occur after two thin active layer portions have merged. In the pinned phase, we can see that the portion of the interface which is inactive is extensive, and appears to extend beyond the width of the active layer gaps. This is a result of interface overhangs, ie that the biofilm fingers are wider at the top than at the bottom, and at the top they may be growing while at the bottom they are not.

### 6.3.2 Linking roughness behaviour and interface pinning

#### Unpinned phase

In this section, I will investigate how the pinning behaviour of the biofilm interface relates to the long-time behaviour of the interface roughness. In Figure 6.9 I show examples of the three distinct types of roughness trajectories that we saw in the last chapter - a smooth steady state for low roughness, a fluctuating steady state

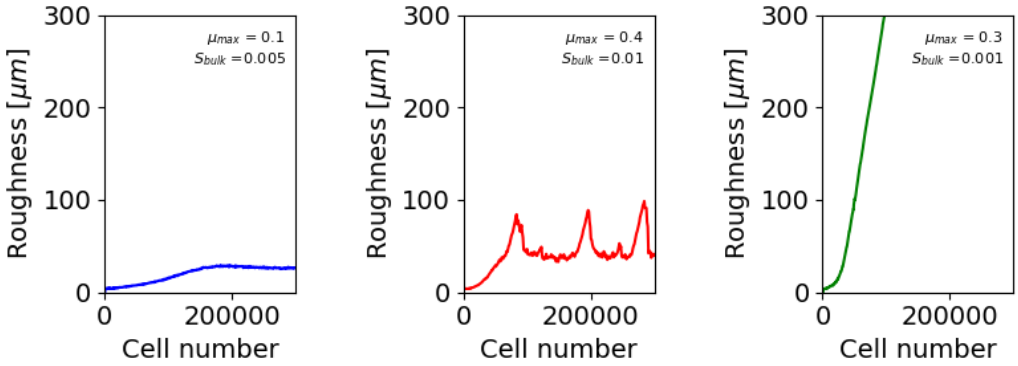



Figure 6.9: Examples of each of the three roughness trajectories identified in the previous chapter. Other than the $\mu_{max}$ and $S_{bulk}$ specified on the plots, the simulation parameters are as in Table 3.1 in Chapter 3.

for intermediate roughness, and a monotonically increasing roughness which can reach very large roughnesses ($\mu_{max} = 0.1$, $S_{bulk} = 0.005$; $\mu_{max} = 0.3$, $S_{bulk} = 0.001$ and $\mu_{max} = 0.4$, $S_{bulk} = 0.01$ and otherwise the same simulation parameters as in Table 3.1 in Chapter 3). We can see that by comparison, the unpinned, depinned and pinned phases we saw in Figure 6.7 correspond to these three roughness behaviours. I discuss in more detail how these pinning behaviours produce the interface behaviour in the rest of this section.

Firstly, I examine the unpinned phase. We know from the previous chapter that the smooth biofilm active layer has a thick average active layer, and that the standard deviaiton of the active layer is small compared with the other phases. We also saw in the snapshots of Figure 6.6 that in this phase there are no discontinuities in the active layer. This means we can understand the constant steady state roughness as being the result of an active layer which is relatively uniform across the interface, meaning the whole interface is moving forward in a similar manner, with few fluctuations along the interface. The reason a steady state appears is a little more complicated, and I discuss this further in the next section.

### Depinned phase

Secondly, I examine the depinned phase. In Figure 6.6 we saw snapshots of the depinned phase. In this phase, pinning occurs, such that a part of the interface gets left behind. This sets the minimum bound on the interface width while the top of the interface, which remains active, continues to grow. As such, the

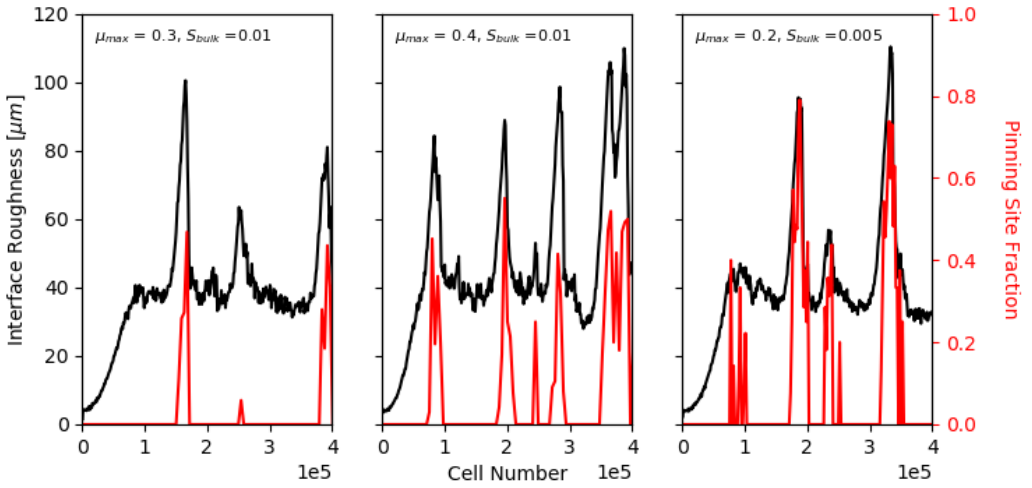



Figure 6.10: This Figure shows the interface roughness plotted together the stationary interface fraction (or in other words the pinned interface fraction) for three simulations which are in the depinned phase (as defined in the text).

difference between the highest and lowest parts of the interface grows, or in other words the interface width, or roughness, grows. When the pinning sites move back together, the distance between the highest and lowest parts of the interface start to decrease, and so the roughness starts to decrease again. This results in the fluctuating roughness behaviour in the depinned phase. The cause of the closing up of the pinning sites is an interesting question, one which I do not have time to investigate in detail but I explore one possible cause in the discussion of this chapter.

I confirm that these pinning-depinning events are responsible for the fluctuations in the interface roughness by considering the fraction of the interface which is stationary over the course of biofilm growth. The stationary parts of the interface are by definition pinned. As we outlined in Section 6.2, the pinned or stationary interface fraction is the number of inactive interface grid squares which have not moved in the previous 6 hours of growth divided by the total number of interface grid squares. Figure 6.10 shows the trajectories of the pinned interface calculation and the interface roughness plotted together. We can see that the peaks in the roughness correspond to the presence of the stationary interface. Since the stationary interface implies the existence of a pinning site, this confirms that the the fluctuations of the interface roughness are due to pinning sites appearing and being overcome as I anticipated in the previous paragraph.

A further interesting feature of the pinning sites in this phase is that they move

around i.e. the pinning sites do not re-pin in the same position each time, as we saw in Figures Figure 6.8 and 6.4. Though we do not have conclusive evidence as to why the gaps in the active layer move around, close inspection of the snapshots in Figure 6.4 gives us some insight. We can see that the arrow A points to the smallest biofilm finger, and arrows B and C track its progression. It appears the two larger biofilm fingers on either side of the small biofilm finger press together until that finger is engulfed. Looking at the shape of the most actively growing cells in these same snapshots (i.e. the parts of the biofilms coloured yellow), we see that the active interface is growing in all directions along the curvature of the interface, including some amount in the horizontal direction. The larger fingers have more growth overall, including in the horizontal direction, and so envelop the smaller fingers. In this manner, the active layer gaps and the pinning sites move around.

Careful inspection of Figures 6.8 and 6.7 suggest the reason that pinning sites appear and are depinned are also to do with the motion of pinning sites. Firstly, I discuss the role of the motion of active layer gaps in the formation of pinning sites. In Figure 6.8, we can see that pinning sites in the depinned phase (ie the red dots of inactive interface) occur after two gaps in the active layer have met, with the larger biofilm fingers engulfing the smaller ones as we suggested above. This suggests that there may be a critical active layer gap size that is needed for a pinning site to form. Secondly, I discuss the role of the motion of active layer gaps in depinning events. The formation of another active layer gap elsewhere in the biofilm seems to happen at a very similar time to a depinning event, for example A and B in Figure 6.8. I think this means that the activity of the active layer is more concentrated in the remaining fingers and so the horizontal 'pushing' is enough to overcome the pinning sites. This suggests that there is a critical active layer gap width (which is different to the gap size needed for a pinning site to form) that cannot be overcome by the horizontal motion of biofilm fingers, leading to the permanent pinning sites we see in the pinned phase.

### Pinned phase

Next, I examine the pinned phase in further detail. We saw earlier that this phase shows a monotonically increasing interface roughness which does not level off, and has pinning sites that do not unpin. Figure 6.11 helps us understand how these two phenomena are related in qualitative terms. We can see that the

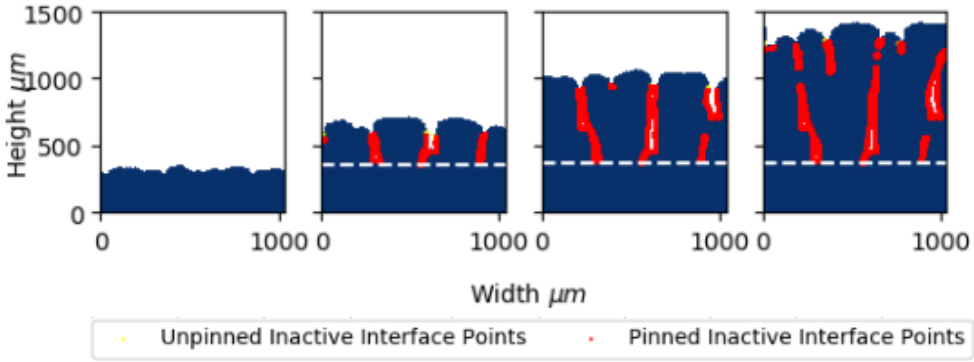



Figure 6.11: This Figure shows snapshots of the pinned phase, and qualitatively demonstrates that the troughs of pinning sites do not move once formed. The red dots are stationary, or pinned interface points. The dottted white line is the position of the minimum interface once pinning sites are formed.

minima of the interface (i.e. the troughs) become inactive and stationary as those parts of the interface lose access to the nutrients. Once these pinning sites have formed, the minima of the interface do not move, and sets the lower bound on the interface width. Meanwhile, the active interface at the top of the biofilm fingers continue to grow, so that the minima and maxima of the interface move away from each other. Since the interface roughness measures the interface width, this implies the roughness is always increasing. I quantitatively confirm this by showing that the interface roughness in this phase increases at the same rate as the speed of the maximum interface for all the simulations in the pinned phase, in Figure 6.12.

### Transitional phase

Finally, I also consider the trajectories we saw in the last chapter that I described as 'transitional' trajectories. These are replotted in Figure 6.13. One of these trajectories reaches a fluctuating steady state before starting to increase monotonically. By analogy with the previous analyses I have made, we can expect that one of the pinning sites did not unpin in this case. The other trajectory seems to reach a smooth steady state before moving to a different behaviour, possibly the oscillating steady state behaviour. We saw in the configuration snapshots of Figure 6.6 that even in the unpinned phase, the active layer is thinner in some positions than others, so it seems likely one of these thinner portions turned into a pinning site. This implies that the phases are unstable, though further

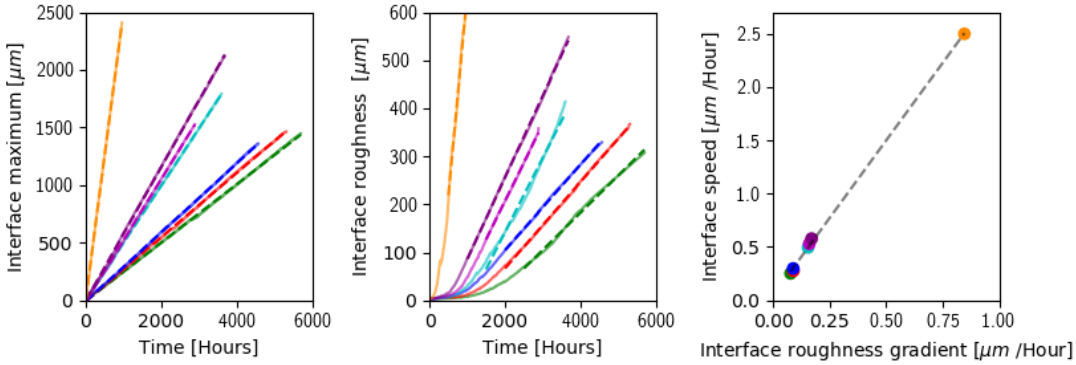



Figure 6.12: This Figure compares the position of the interface maximum and the interface roughness for simulations in the pinned phase, as defined in the text. The left panel of shows the maximum interface speed against time. The central panel shows the interface roughness against time. The right hand panel shows the gradient of the linear portion of the interface roughness against the interface speed (the gradient of the maximum interface). In each of the left most panels, the dashed lines are used to measure the gradients plotted in the right hand panel.

investigations are needed.

### 6.3.3 Linking the Interface Roughness and Active Layer standard deviation

I now examine the relationship between the standard deviation of the active layer thickness and the interface roughness, or in other words the relationship between the fluctuations in the interface position and the fluctuations in the active layer

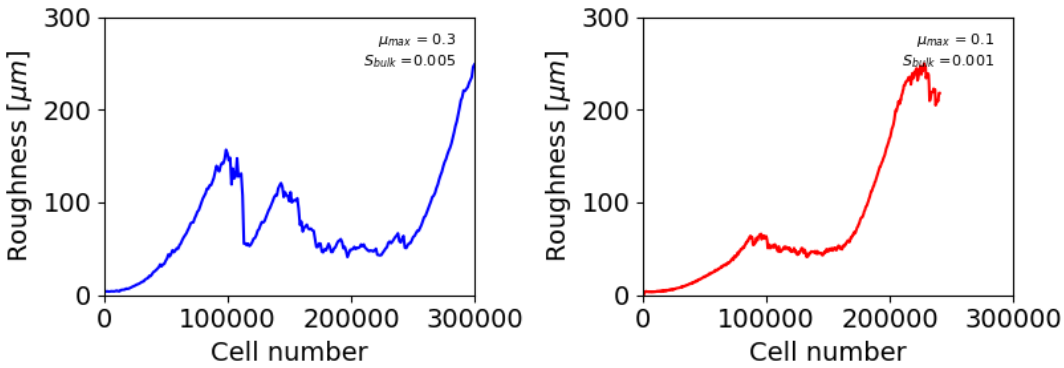



Figure 6.13: This Figure replots in the interface roughness trajectory for the 'transitional' phases described in the text.

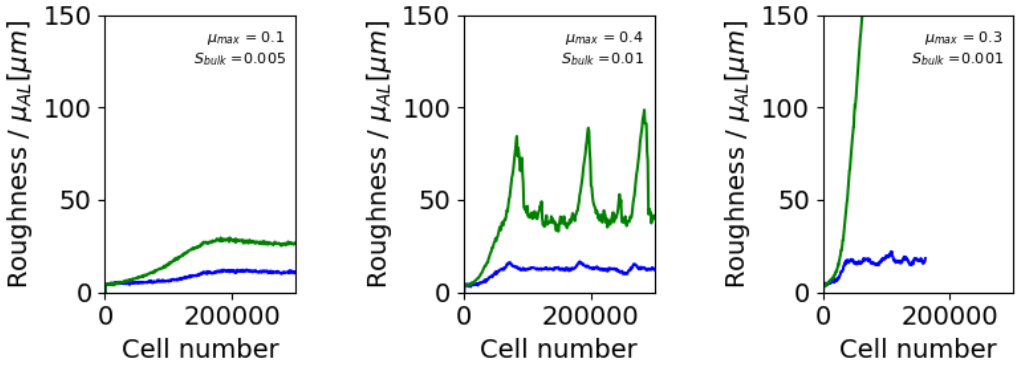



Figure 6.14: This plot shows the trajectories of the interface roughness and the standard deviation of the active layer thickness ($\mu_{Al}$ plotted together. $S_{bulk}$ and $\mu_{max}$ values are as specified in the plot and all other parameters are specified in Table 3.1.

thickness. I first recall from Figure 5.14 in Chapter 5 that the standard deviation of the active layer thickness reaches a steady state in all phases, even when the interface roughness continues to increase. I then plot the roughness with the standard deviation of the active layer for examples of each of the phases in Figure 6.14 and observe that the standard deviation of the active layer thickness steady state is reached at approximately the same cell number as the roughness steady state in the unpinned and depinned phases. This suggests that the fluctuations in the interface position are closely related to the fluctuations in the active layer fluctuations.

In Figure 6.15 I now plot the interface roughness against the active layer standard deviation for the same simulations with varying $\mu_{max}$ and $S_{bulk}$ I have previously examined. The colour coding is such that the lines are plotted in blue at times the standard deviation of the active layer thickness has not reached a steady state, and in green at times when a smooth or fluctuating steady state has been reached. I firstly observe there is a linear relationship between the roughness and the active layer standard deviation in all the plots before the standard deviation of the active layer thickness is reached (i.e. the blue parts of the lines). This indicates their behaviours are initially correlated, suggesting that fluctuations in the active layer are linked to fluctuations in the interface position and visa versa. I secondly observe that after the standard deviation of the active layer thickness has reached a steady state (i.e. the green parts of the lines), the relationship between the roughness and the standard deviation of the active layer thickness often changes and can become uncorrelated.

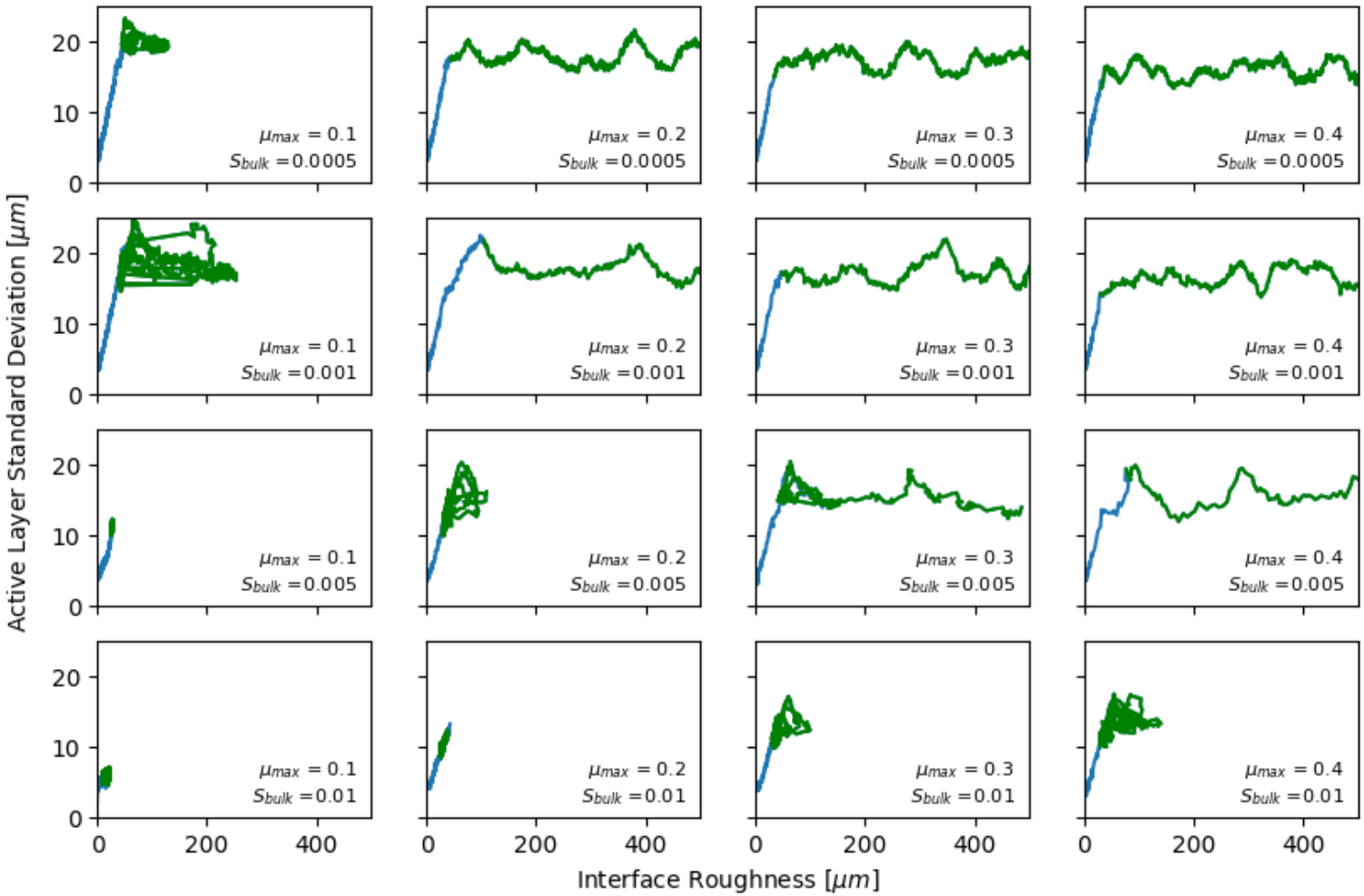



Figure 6.15: This Figure shows the Interface roughness plotted against the standard deviation of the active layer thickness for varying maximum specific growth rate $\mu_{max}$ and maximum nutrient concentration $S_{bulk}$. The colour coding is such that the lines are plotted in blue at times the standard deviation of the active layer thickness has not reached a steady state, and in green at times when a smooth or fluctuating steady state has been reached.

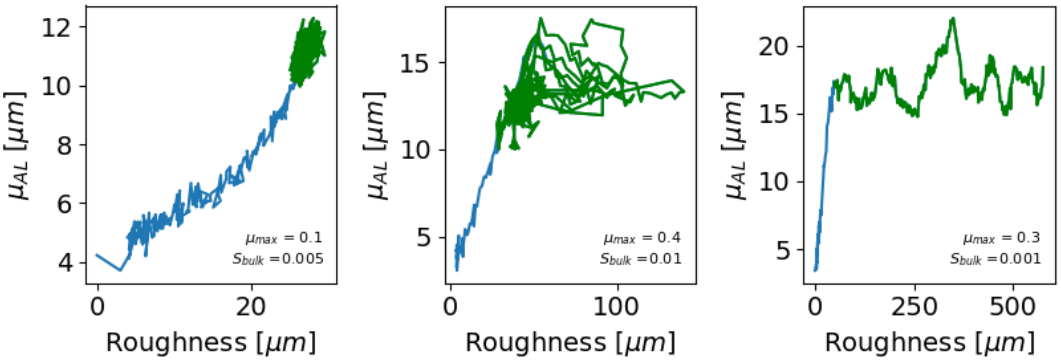



Figure 6.16: This Figure shows three examples of the interface roughness plotted against the standard deviation of the active layer thickness that are representative of the three phases discused in the text. The left panel is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.1$, the central panel is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.4$ and the right panel is for parameters $S_{bulk} = 0.0005$, $\mu_{max} = 0.4$.

I observe three distinct behaviours after the standard deviation of the active layer thickness has levelled off in Figure 6.15, which I highlight in Figure 6.16. Each of these examples corresponds to each of the three phases we have discussed previously. The left panel shows the unpinned phase, in which I observe that after the steady state is reached, the roughness and the active layer standard deviation of the active layer remain close to their value at the point at which the steady state is reached, indicating that they are still correlated at this point. The right hand panel shows the pinned phase, where we can see the fluctuations in the interface position and the active layer become uncorrelated, with the standard deviation of the active layer remaining roughly constant as the roughness continues to increase. The central panel shows the depinned phase, there seems to be a mix of these two behaviours. Thus implies that gaps in the active layer cause the interface fluctuations to become uncorrelated from the flcutuations in the interface, as the interface is then able to change position in this region without changing the distribution of the active layer. Together this suggests one way of understanding pinning is that the size of the active layer fluctuations becomes larger than the width of the active layer, allowing the fluctuations in the interface position to become uncorrelated from the active layer behaviour.

## 6.4 Discussion

In this chapter I have elucidated the diversity of biofilm interface roughness behaviour by examining the behaviour of the active layer and in particular the phenomenon of interface pinning. I found that the three qualitatively distinct roughness behaviours identified in Chapter 5 each correspond to a distinct type of pinning behaviour. Firstly, I found an unpinned phase whose interface roughness remained low and reached a constant steady state. I also found a phase where pinning sites appeared but were overcome which I dubbed the depinned phase, which corresponded with a fluctuating long-time roughness behaviour. Finally, I found a pinned phase in which pinning sites remained once formed and the interface roughness continues to increase with time. As we saw in Chapter 2, the most detailed studies of the interface roughness of biofilms and colonies has come from interface growth theory, and therefore some of the clearest comparisons to our work can be drawn from this literature. There is nevertheless also much in my work that is relevant to previous work in reaction-diffusion modelling of biofilms, which we also discuss here.

I firstly consider the resemblance of our three phases to the pinning-depinning transitions discussed in interface growth theory. We recall from Chapter 2 that a pinning-depinning transition involves qualitatively the same behaviour as I observe - a flat phase with no pinning sites, a pinned phase and a intermediate depinned phase in which pinning sites are overcome [33, 34]. Indeed this literature inspired the names of my phases. However, the quantitative connection between my work and interface growth theory is less clear. In common with some others in the literature, I observe a monotonically increasing roughness in the pinned phase, which is not consistent with the pinned phase in the simplest version of a pinning-depinning transition (the quenched KPZ framework which we examined in Chapter 2) which requires a steady state in order to measure the defining scaling exponents [33, 99, 104]. It is possible that my pinned phase is the Diffusion Limited Aggregation Phase, as it is reasonably well established in the interface growth theory and reaction-diffusion literatures that the roughest biofilms tend to be diffusion dominated, though it is not completely clear if this is the case in all models. Though further work would be needed, including measurement of scaling exponents, my depinned phase could be consistent with the quenched KPZ phase.

In understanding the active layer dynamics, I also believe it is important to

consider the role of the mechanical interactions. In the results section above, I explained how the growth of the cells in the horizontal as well as vertical directions meant that active layer gaps moved around and were responsible for closing up pinning sites in the depinned phase. Specifically, the fact that the cells grow in all directions, not just in the direction of the nutrients, means that larger biofilm fingers 'push' the smaller ones around. How far cells move apart from each other and in what direction when they are growing is determined by the shoving algorithm in iDynoMiCS, which we outlines in Chapter 3. I therefore think we need to understand the active layer dynamics as a combination of the response to the local nutrient field and the mechanical interactions. The cells which are in the active layer have access to nutrients and are therefore able to grow. In the iDynoMiCS simulations, the cells which are mechanically interacting are the cells which are growing, and therefore the cells in the active layer. I therefore think it would be interesting to see how the phase behaviour might change if the mechanical interactions were changed - for example, I think that stronger mechanical interactions might mean that larger pinning sites can be closed up and so lower nutrient concentrations would be needed in order to see the pinned phase. The role of mechanical interactions in phase transitions of bacterial communities [14, 16], but I think a careful study of how the mechanical interactions influence the active layer dynamics would be fruitful. I note, however, that iDynoMiCS does not model the mechanical interactions in a physically realistic manner, and so may not be the best software for proceeding with these investigations.

As I said above, the depinning phase that I identify in my simulations is is similar to the quenched KPZ phase. An important question if this is the case is the source of the quenched noise. As we saw in Chapter 2, the primary source of quenched noise that has been studied in previous work is inhomogeneities in the medium [34, 80]. Clearly, in my simulations there are no inhomogeneities in the medium. Bonachela *et. al* also observe a quenched KPZ type phase and suggest that nutrient limitation could provide a source of quenched noise in biofilm growth [21]. However, I believe that it is more correct to understand the quenched noise as gaps in the active layer, whose behaviour is influenced by mechanical interactions and nutrient limitation as I have argued above. Additionally, as we saw in Chapter 2, there are models other than the quenched KPZ equation which involve a pinning-depinning transition. Since the source of the pinning sites appears to be intrinsic to biofilm growth, I also think it would be interesting to investigate whether self-organised pinning might be a more appropriate class of growth then the quenched KPZ class [34].

While I believe it is in principle possible to formally classify each of the three phases I see in my work using the framework of interface growth theory, I chose not to proceed with this avenue of inquiry for several reasons. Firstly, performing enough repeat simulations to get good enough statistics, e.g. to calculate scaling exponents, was not computationally feasible. Secondly, though models and experiments being in the same universality class ensures the same physical and biological processes are causing the growth behaviour, it is not often clearly known which processes correspond to which growth class [47]. I instead proceed in the next chapter with an order parameter based description of the three phases.

## 6.5 Summary

In this chapter, I examined in detail the dynamics of the active layer and of interface pinning. I found three distinct behaviours - no pinning, pinning sites that appear and can be overcome, and pinning sites which remain over the course of biofilm growth once they have formed. I explained how the three distinct long time interface roughness behaviours we saw in the last chapter are related to these distinct regimes of pinning. I also explained how we can understand the difference between biofilms that have a steady state roughness and those which do not. I then discussed the relationship of the phases I have found to those identified in the interface growth theory literature.

# Chapter 7

# Analysing the biofilm phase behaviour.

## 7.1 Outline

In the previous chapter I found three distinct phases in the long time biofilm growth behaviour, each with distinct interface roughness trajectories which are closely related to the behaviour of the active layer. I also examined in detail an interface pinning phenomenon, in which gaps in the active layer caused parts of the interface to become stationary, or pinned, and therefore to lag behind the moving front. The interface pinning phenomenon allowed us to understand the origin of the three phases of biofilm growth behaviour. Specifically, the three phases could be classified as an unpinned phase, in which the active layer remains continuous throughout the biofilm growth and no pinning sites arise, a depinned phase in which pinning sites arise but are eventually overcome and a pinned phase in which pinning sites arise and remained to long times. We also saw in Chapter 5 how each of these phases had a distinctive type of interface roughness trajectory - a smooth steady state for the unpinned phase, a fluctuating steady state for the depinned phase and a monotonically increasing roughness for the pinned phase. In this chapter, I build on these insights to attempt to build a phase diagram for biofilm growth.

Specifically, in this chapter I analyse biofilm growth in terms of a control parameter and an order parameter. This means I aim to plot a suitable order

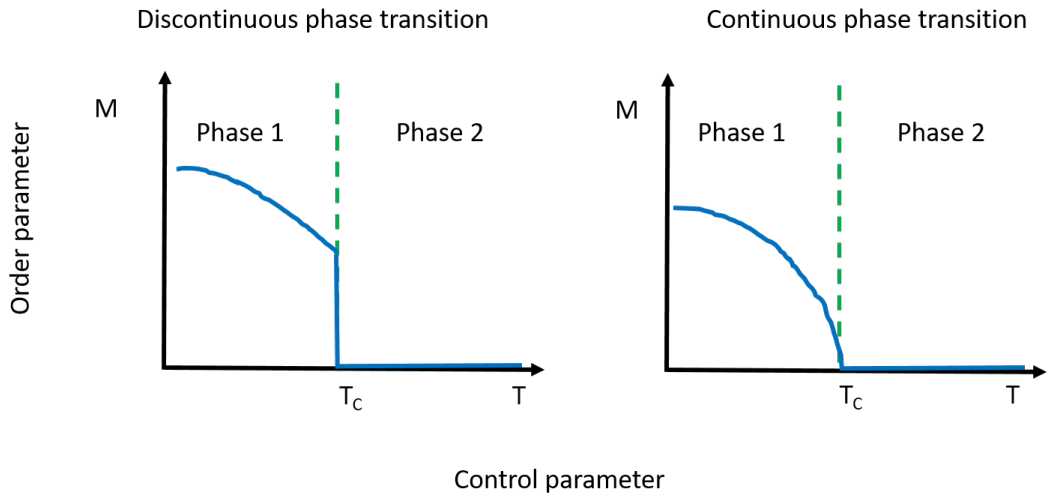



Figure 7.1: Sketches of the two main kinds of phase transitions for the example of magnetic materials, where the temperature $T$ is the control parameter and the magnetisation $M$ is the order parameter. The transition occurs at a critical temperature $T_C$. The left hand panel shows a first-order phase transition, where the order parameter is discontinuous at the point of transition and the right hand panel shows a second-order phase transition which is continuous at the point of transition.

parameter (y-axis) as a function of a control parameter (x-axis), such that when the control parameter is varied, the order parameter undergoes a transition. In other words, the order parameter distinguishes between the different phases. We can see a sketch of what this might look like in Figure 7.1, for the classic example of phase transitions in magnetism, where the magnetisation is the order parameter and the control parameter is the temperature. Figure 7.1 illustrates two distinct types of phase transition - first order and second order transitions, in which the order parameter behaves discontinuously and continuously at the point of the transition, respectively. Continuous and discontinuous phase transitions are also characterised by different phase transition kinetics. Let us suppose that the order parameter undergoes a sudden shift such that the system will transition from one phase to another. In the case of a first order transition, the system requires a critical fluctuation to overcome a free energy barrier and nucleate the new phase, which is a local process, while in the case of a second order transition, the new phase appears spontaneously, everywhere in the system, as there is no free energy barrier to be overcome.

In this chapter, my main aim is to find suitable control and order parameters to allow a phase diagram like those in Figure 7.1 to be plotted. The motivation for

finding the control and order parameters is to define the physical nature of the transition, and ultimately, by understanding it better, to predict or control the transition. It is also important for understanding which quantities have a defining influence on the system behaviour. I note that in this sense, the control parameter does not necessarily have to be 'controllable' in a simulation or experimental setting to be useful.

I begin by outlining my suggested order parameter. I look at the quantities which I have previously found to be key for the phase-defining pinning behaviour - the inactive and stationary parts of the interface. I then go on to investigate potential control parameters, by looking to see which of the quantities I defined in the previous chapters - the active layer thickness and the standard deviation of the active layer thickness - might distinguish between the three phases when used as a control parameter. I outline evidence that suggests that the normalised standard deviation of the active layer thickness may be a better control parameter than the average active layer thickness. As I will discuss, in practice this work involved investigating a much greater number of potential control and order parameters, and I present here a condensed version of this investigation for clarity.

## 7.2 Results

### 7.2.1 Order parameter

I begin by looking for an order parameter, which I wish to be a distinguishing parameter for the three phases. I therefore begin by recalling the important distinguishing feature of the three phases which we saw in the previous chapters - the presence of inactive and stationary parts of the interface. Figure 7.2 shows (again) snapshots of the three phases with the inactive and stationary parts of the interface coloured in yellow and red respectively. We can see that in the unpinned phase the interface does not become inactive or stationary and remains smooth. In the depinned phase, inactive and stationary parts of the interface, or pinning sites, can arise but are overcome. In the pinned phase, inactive and stationary regions constitute a significant fraction of the interface as pinning sites are formed and are not overcome.

I now examine the fractions of the interface which are inactive and stationary, calculated as described in Chapter 6. Figure 7.3 shows us the inactive and

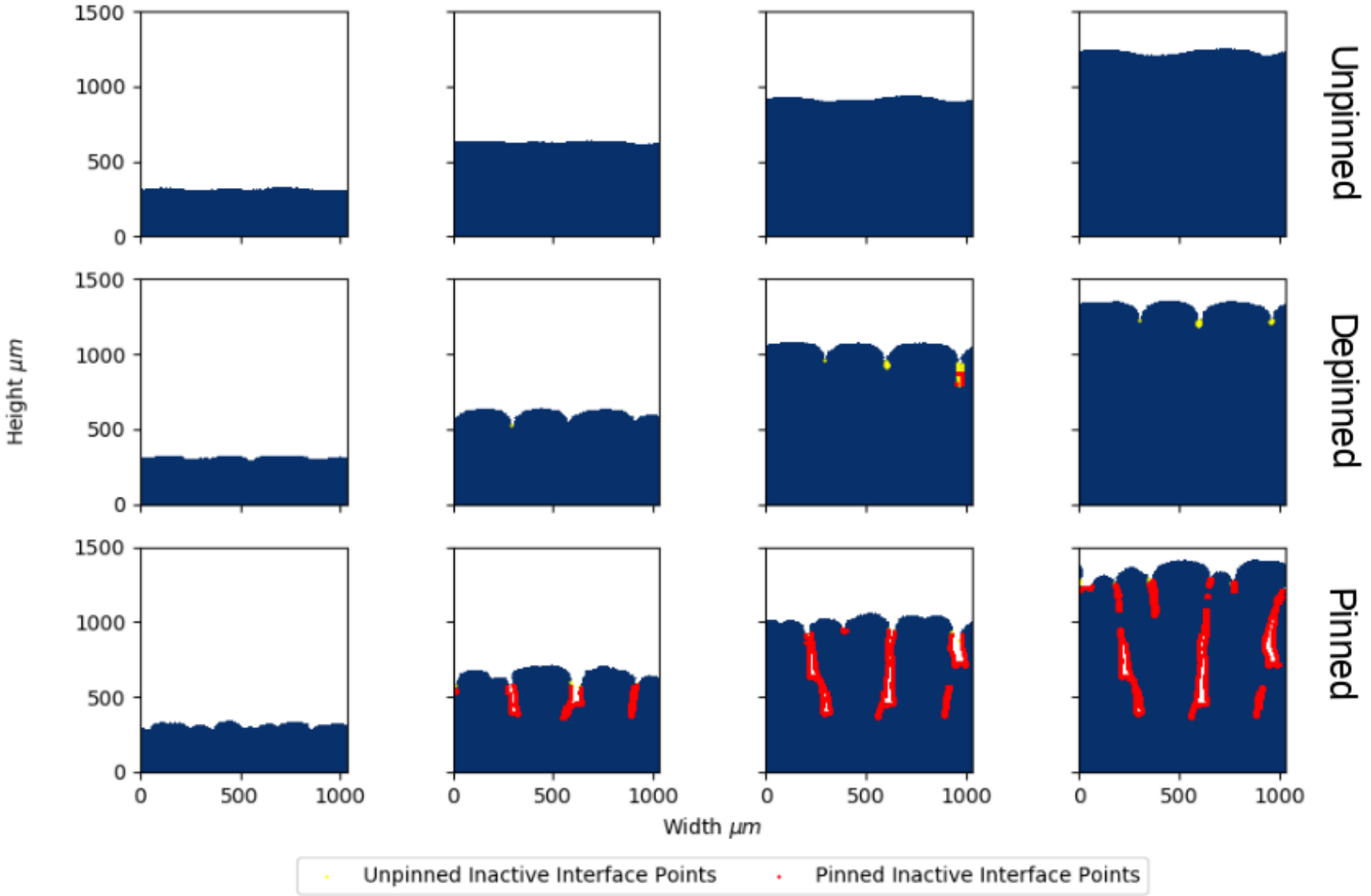



Figure 7.2: This Figure shows snapshots of the biofilm configurations for simulations that represent each of the three phases. The yellow dots are parts of the interface which are inactive. The red dots are parts of the interface that are inactive and stationary (or in other words pinned as I define in the text). The top row of snapshots is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.1$ which represent the unpinned phase described in the text. The central row of snapshots is for parameters $S_{bulk} = 0.01$, $\mu_{max} = 0.4$ which I refer to as the depinned phase in the text. The bottom row of snapshots is for parameters $S_{bulk} = 0.0005$, $\mu_{max} = 0.4$ and represents the pinned phase. Each column of snapshots is for biofilms with 25000, 50000, 75000, 100000 cells.

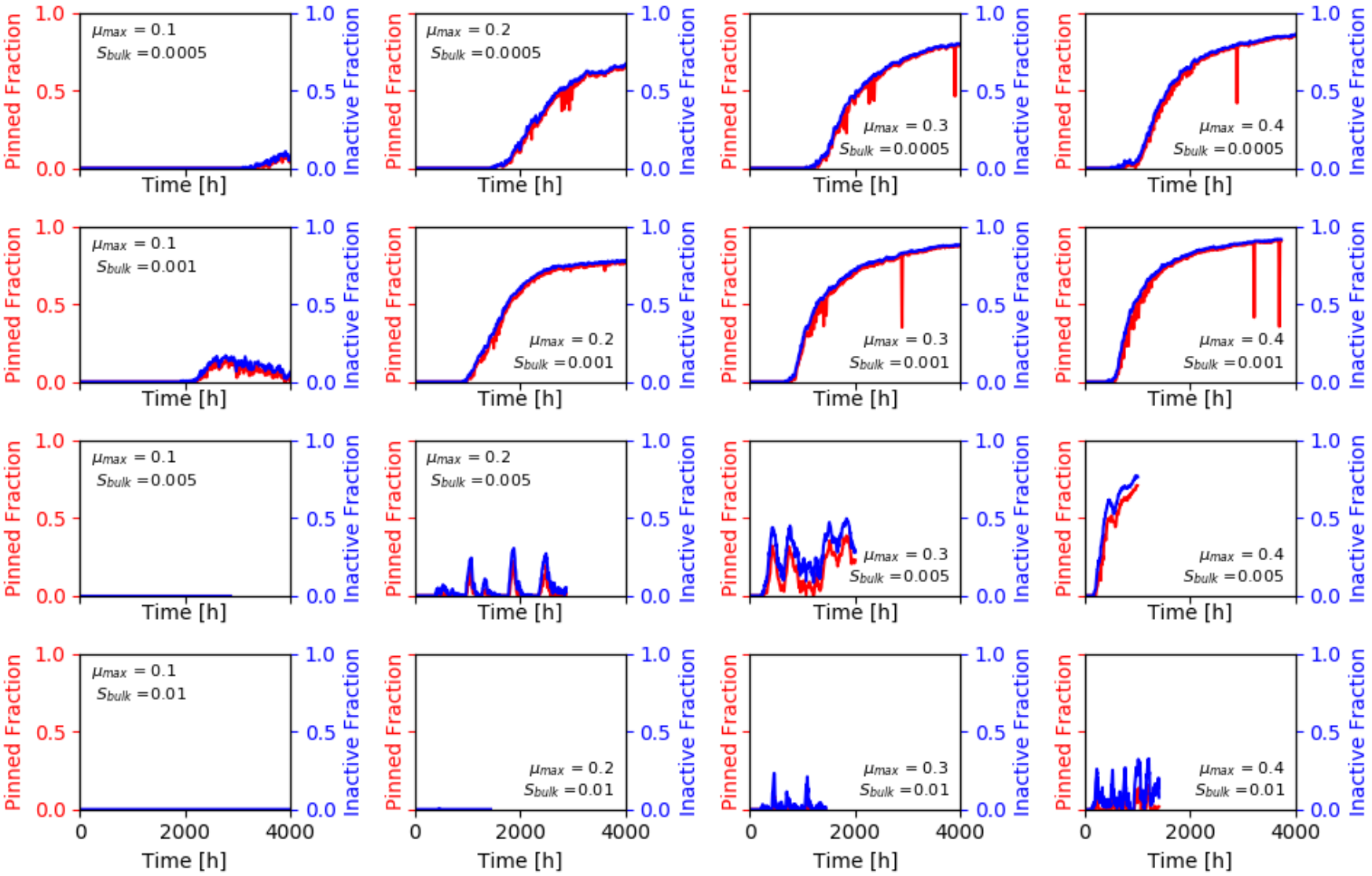



Figure 7.3: This figure shows the trajectories of the inactive interface fraction (blue) and the pinned interface fraction (red) for varying $\mu_{max}$ and $S_{bulk}$. In several of the subplots, most notably $\mu_{max} = 0.4, S_{bulk} = 0.01$ and $\mu_{max} = 0.4, S_{bulk} = 0.005$, there are several noisy events in the pinned interface fraction. I believe these are from the way the pinned interface fraction is calculated, rather than being events in the simulated biofilm. There was not sufficient time to investigate fully, but I think this is a result of iDynoMiCS output files not being produced at 6 hour intervals at some clipping transitions, which is necessary for the pinned interface calculation.

stationary interface fraction trajectories for the grid of simulations with varying $S_{bulk}$ and $\mu_{max}$ which we examined in the previous chapters. I first make an observation that is common to all these graphs. In all cases where the inactive and stationary fractions are non-zero, the inactive fraction is a larger portion of the interface, though it often has a similar trajectory. This can be understood from the snapshots of Figure 7.2. We saw that the gaps in the interface appeared, or in other words the interface became inactive, which then often led to the interface becoming stationary as it was no longer growing. Active layer gaps, or inactive interface sections, are often a little wider than the pinning sites (stationary interface sections) themselves, leading to a situation where the inactive fraction of the interface is often slightly larger than the stationary interface fraction.

I now discuss the types of trajectory form shown by the inactive and stationary interface fractions. We firstly see that for large $S_{bulk}$ and small $\mu_{max}$, both the inactive and stationary interface fractions are zero. This corresponds to the unpinned phase, which we have already seen (eg in Figure 7.2) does not have any active layer gaps or pinning sites. Secondly, we see that for small $S_{bulk}$ and large $\mu_{max}$ (the pinned phase), the inactive and pinned interface fractions level off at values of 0.7-0.9. We can also understand this by considering the shapshot images in Figure 7.2, in which the pinning sites fix the minima of the interface while the growing front continues to advance. This produces an increasing length of inactive interface, while the length of the growing front remains approximately the same length. At intermediate $S_{bulk}$ and $\mu_{max}$ (the depinned phase), the trajectories show intermittent periods during which the inactive/stationary interface fraction is either zero or non-zero; these correspond to the transient pinning events described in Chapter 6. There are also two of the trajectories ($S_{bulk} = 0.005$, $\mu_{max} = 0.3$; $S_{bulk} = 0.001$, $\mu_{max} = 0.1$) which do not fall into these categories. These are simulations which undergo transitions between two different phases of biofilm growth.

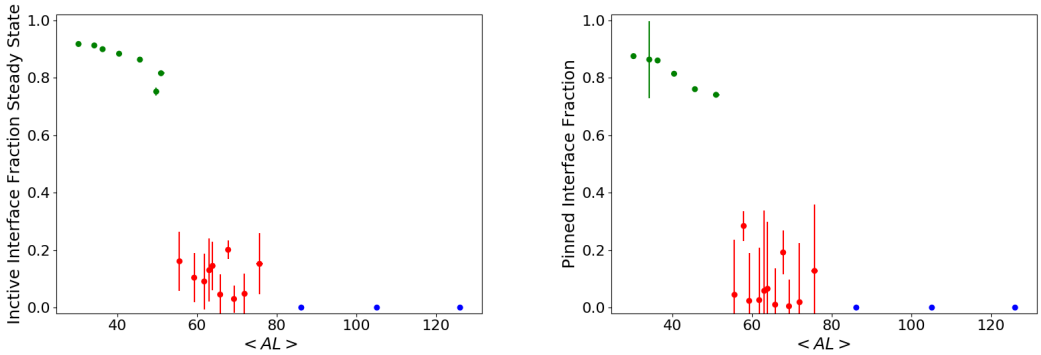
In the rest of this chapter, I choose to use the steady-state averages of the inactive interface fraction and the pinned interface fraction as potential order parameters. Though longer runs of some of these simulations would be ideal, this was not computationally feasible within the time frame of this project to continue these simulations. As I detailed in Chapter 4, it is not possible to clip a significant number of cells from the very rough biofilms without removing portions of the interface, meaning it is not possible to speed up these simulations any further by removing cells from the simulations. Nevertheless, these averages do serve the

important function of any order parameter by clearly distinguishing the different phases i.e. each of the phases has values of these quantities in ranges which do not overlap with one another. As previously mentioned, the inactive interface fraction and pinned interface fraction remain at zero. The simulations previously identified as being in the depinned and pinned phases have ranges of the average steady state inactive interface fraction of 0.014-0.161 and 0.752-0.912 respectively, and ranges of the average steady state pinned interface fraction of 0.010-0.284 and 0.741-0.875 respectively. I will discuss in the next section, after I have compared these two potential order parameters with potential control parameters.

### 7.2.2 Control parameter

In the previous section, I found potential order parameters in the steady state inactive and pinned interface fractions and I now investigate potential control parameters. I recall from the introduction of this chapter that when the control parameter is varied, the order parameter undergoes a transition, as we saw some examples of in Figure 7.1. In our case this would mean a quantity that, when smoothly varied, tunes the system between the three phases i.e. the unpinned, depinned and pinned. We recall from Chapter 2 that Nadell *et. al.* suggested the active layer thickness as a combined parameter that controls the behaviour of the system [28]. I therefore use the the average active layer thickness as an initial candidate for a control parameter. However we also saw in the previous chapters that the standard deviation of the active layer thickness reaches a steady state and that this also includes dynamical information on the active layer fluctuations. Therefore the standard deviation of the active layer thickness is also a candidate for the control parameter. In the rest of this subsection I look to find a control parameter which best predicts the behaviour of the average inactive and pinned interface fractions. I do this by looking for a quantity which produces a phase diagram like that shown in Figure 7.1 - i.e. the data collapse onto a single line when plotted in terms of the control parameter and our chosen order parameter, with the three phases being clearly distinguished since they occur over distinct ranges of the control parameter. In practice I tried this for many quantities I do not present here, including the interface speed and the total activity. For clarity, I omit much of that work here. I also performed additional simulations beyond those we saw in Figure 7.3 with $S_{bulk}$ and $\mu_{max}$ values intermediate to what is in the grid of values previously presented.

I begin by considering the average active layer thickness as a candidate control parameter. In Figure 7.4 we can see our chosen order parameters, the inactive interface fraction and the pinned interface fraction averages, plotted against the average active layer thickness, with the different phases we defined in the previous chapters coloured accordingly. We can see that the the different phases are well distinguished (i.e. they are not all mixed in with one another) here. Though there is not a perfect collapse onto the a single line, we can see that there is a general trend that a thicker active layer thickness correlates with a transition between the unpinned to depinned and depinned to pinned. We can see that as we anticipated earlier, the unpinned phase has the thickest active layer and zero inactive and pinned fractions. The pinned phase has the thinnest active layer thickness, and a highest inactive and pinned interface fractions. There is an intermediate regime for the depinned phase, but it is apparent from the lack of data collapse that the average active layer thickness is not the only factor that determines the behaviour here.




(a) Inactive Fraction

(b) Pinned Fraction

Figure 7.4: This Figure shows the average active layer thickness plotted against the inactive interface fraction in the left panel and the pinned (or stationary) interface in the right hand panel.

Motivated by the poor data collapse in Figure 7.4, I now investigate whether including the standard deviation of the active layer thickness in the control parameter can improve the data collapse. I now investigate the potential use of the normalised steady state active layer thickness standard deviation. I argued in Chapter 6 that the active layer thickness fluctuations as well as the average active layer thickness were important in understanding the different interface pinning behaviour in the three phases. These active layer thickness fluctuations can be described by the standard deviation of the active layer thickness, which we have seen in Chapter 5 reaches a steady state in all phases, including in the pinned

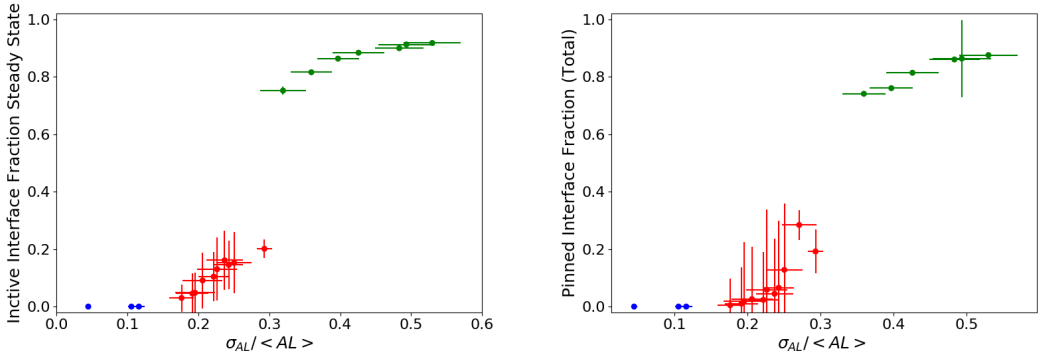
phase where the roughness does not reach a steady state. I choose the normalised standard deviation of the active layer thickness, i.e. the steady state active layer thickness standard deviation divided by the average active layer thickness for each of the simulations to capture the size of the fluctuations relative of the overall thickness.

In Figure 7.5 we can see the inactive and stationary interface fractions plotted against the normalised standard deviation of the active layer thickness. We can see that this provides a better data collapse than the average active layer thickness alone, and increasing the normalised active layer thickness standard deviation moves the phase from the unpinned phase through depinned to pinned. This implies that to transition between the unpinned and pinned phase we need not only decreasing active layer thickness but also increasing active layer thickness standard deviation, or in other words increasing active layer thickness fluctuations. We saw in Chapter 6 that while the average active layer thickness can remain constant, the local active layer thickness can fluctuate dramatically and produce a variety of complex dynamics. Therefore, Figure 7.5 suggests an important role for the active layer dynamics in the physics of these phases, not just the average active layer thickness as discussed in Nadell *et. al.* and others [16, 28, 65].

Finally, I consider whether the inactive interface fraction or the pinned interface fraction is the better order parameter. We can see in Figure 7.5 that the phase diagrams look quite similar for both order parameters. However, the inactive interface fraction seems to produce a slightly better collapse of the data, and has smaller error bars. Further work would be needed to establish whether these small differences are significant, but in the absence of significant differences between the two parameters, I suggest the pinned interface fraction is the more intuitive order parameter. This is because in our analysis, we have seen the key distinguishing feature between the phases is the pinning behaviour, and therefore having an order parameter which measures this behaviour directly is more intuitive.

### 7.2.3 Discussion

In this chapter I have considered the phase behaviour discussed in the previous chapters from the perspective of constructing a phase diagram. Specifically, I have attempted to find the order and control parameters of the transition, to produce a phase diagram of the form shown in Figure 7.1. I investigated many potential




(a) Inactive Fraction (b) Pinned Fraction

Figure 7.5: This Figure shows the normalised standard deviation of the active layer thickness (the potential control parameter) plotted against the inactive interface fraction in the left panel and the pinned (or stationary) interface in the right hand panel (the potential order parameters).

parameters, but presented here only those which gave the most promising results. I argued that it is possible to distinguish the different phases using the steady state fraction of the inactive interface and the steady state fraction of the interface which is stationary (or in other words pinned), and so these are potential order parameters. I then investigated a number of quantities as the potential control parameter, and I tested which of these was best able to predict the behaviour of the pinned interface fraction and so produced the best collapse of the data onto a single line. On this basis, I argue that the normalised standard deviation of the active layer thickness, rather than the average active layer thickness alone, is the better control parameter, indicating that the active layer dynamics are important in driving the physics of the phase behaviour as well as the average active layer thickness behaviour. What I have here is not a traditional control parameter; rather, it could be called a 'dynamical control parameter' as it is describing the long-time steady state behaviour.

As mentioned in Section 7.1, a phase diagram (order parameter v control parameter) has two main functions. Firstly, it can be used to predict which phase will be produced from the starting set up and second, it serves as a tool to more deeply understand the physics of the system. However, the phase diagram I have produced in Figure 7.5 is of limited use in predicting what will happen in a given simulation. This is because our control parameter involves the standard deviation of the active layer thickness that we do not know how to link to the input parameters of the system, though I will discuss further work which may make

this possible later in this section. Additionally, the control and order parameters involve the steady state of biofilm growth, which means these properties could not be measured at early times to predict the distinctive steady state behaviours, as would be possible if the control parameter was the average active layer thickness alone. Nevertheless, my proposed control parameter can still be used to help understand the physics of what is going on.

A key aspect of what I have found is that the standard deviation of the active layer thickness (i.e the dynamics of the active layer) matter to the phase behaviour as well as the average active layer thickness. As I have mentioned, this is distinct from the work of several authors, including Nadell *et. al.* [28], who propose that the average active layer thickness alone can predict the spatial structure of the biofilm. My work implies that two biofilms with the same average active layer thickness could have different active layer dynamics, and while clearly some of the same influences are at play in both the average and standard deviation of the active layer thickness, there are some distinct processes occurring.

Additionally, the importance of active layer dynamics in the phase behaviour also implies the importance of mechanical interactions. I argued in the discussion of Chapter 6 that the average active layer thickness was likely determined by parameters describing the diffusive behaviour alone, while the active layer dynamics could be explained only by considering mechanical interactions as well as the diffusion of nutrients. I argued that these mechanical interactions were particularly important in closing up the pinning sites in the depinned phase, which in Figure 7.4 we see is the phase that is least well described by the average active layer thickness alone. This leads me to suggest that one might be able to find a good control parameter consisting of some combination of the input parameters of the system - but that it would likely need to include mechanical parameters as well as diffusive ones. As I have discussed previously, iDynoMiCS represents mechanical interactions only in a crude way, meaning that to further investigate the role of mechanical interactions in a possible control parameter one might need to use a simulation algorithm that deals with mechanical interactions more rigorously, for example the recently released software by Jayathilake *et. al.* [105].

Another aspect of the physics which it is possible to examine with the help of Figure 7.5 is whether our system shows a first or second order transition (i.e. whether the transition is discontinuous or continuous, as previously shown in in Figure 7.1). At first sight, it appears that in our system, the transition between

unpinned and depinned phases is continuous, as the data smoothly transitions between the two phases when plotted against the inactive interface fraction and the pinned interface fraction. For the depinned to pinned transition, the data is less conclusive, though the large gap between the range of values of the inactive and pinned interface in the pinned and depinned phases suggests it is possibly discontinuous. It is also possible that we actually have only one transition occurring, from unpinned to pinned, and that this transition is discontinuous, with the depinned and early pinned phase being finite-size effects i.e. that in an infinite system the depinned phase would not exist. We recall that the occurrence of a first-order transition requires a critical fluctuation to overcome a barrier. This picture of a discontinuous transition between depinned and pinned phases, smoothed out by finite size effects, is consistent with my earlier work where I suggested that there may be a critical width of a pinning site to allow closure by the growth of adjacent biofilm fingers. Interestingly, if our system does show a discontinuous transition between unpinned and pinned phases this would provide a contrast to the classic work of Dockery and Clapper on fingering instabilities in biofilm growth, since their work suggested that, due to the inhomogeneous nutrient field, the interface is always unstable to perturbations of any size and so the transition to a rough biofilm is by implication a continuous transition. Their work did not, however, take account of mechanical interactions which my work suggests may be important in the closure of pinning sites.

While my work has clearly shown that there are three distinct phases of behaviour, further work is needed to confirm the best way of classifying them. Most obviously, simulations which vary a wider range of starting parameters than $S_{bulk}$ and $\mu_{max}$ would be necessary to confirm the order and control parameter combination which I argue for in this chapter. It may also be useful to extend the length of the simulations for those simulations which do not fully reach a steady state for the inactive interface fraction (e.g for simulation with $\mu_{max} = 0.4$, $S_{bulk} = 0.005$), though as this is an extremely computationally intensive task. It would also be interesting to investigate different system sizes, which would mean we could confirm whether system size effects are playing a role.

### Summary

In summary, this chapter has attempted to quantify the different phases identified in the previous chapters. While further investigations would be required to

confirm this, I argue that the normalised steady state of the standard deviation of the active layer thickness is the most appropriate control parameter, Thus, I define a 'dynamical control parameter' which tells us about the steady state of the biofilm behaviour and indicates a key role for the active layer dynamics as well as the average active layer thickness.

# Chapter 8

# The genetic spatial structure of a growing biofilm.

## 8.1 Outline

In the previous chapters of this thesis, I have established three distinct phases in biofilm morphology, or macroscopic spatial structure. These are the flat, depinned and pinned phases and we have shown they each have distinctive active layer and interface roughness behaviour. In this chapter, I build directly on this to examine the arrangement of different lineages of cells within the biofilm, or genetic spatial structure, for each of these phases. By doing this, I hope to gain a better understanding of the relationship between the biofilm morphology and the genetic spatial structure.

As we discussed in Chapter 2, the genetic spatial structure of biofilms has a strong influence on the resulting evolutionary behaviour by determining which genetic variants are available to respond to selective pressures. Within spatially constrained biofilms, the cell lineages which come to prominence are not necessarily those cells with the highest fitness, as the cells which have access to nutrients and hence expansion paths are influenced by the same processes which determine the spatial structure which we looked at previously [35, 82]. For this reason, it is relevant to study the genetic diversity which results in spatially structured populations in the neutral case, where all cells are, *a priori*, equally fit, in order to better understand the influences on selection within biofilms that

result from spatial structure rather than cell fitness.

In this chapter, I investigate a number of different measures of genetic diversity, and compare their outcomes for each of the different phases discussed in this thesis. I begin by tracking the progeny of the different initial cells, or the cell families, and look at how diversity is lost from the active layer as biofilm growth progresses. I find that pinning sites act as important sites of diversity loss. I then use a measure of genetic diversity which involves tracking the position of the cells in the genetic tree using the genealogy number. This allows me to investigate more fully the changes in which cell lineages are present in the active layer as the biofilm grows, rather than just the loss of initial cell lineages. I finish the chapter by discussing the contribution of this work to the literature on biofilm genetic spatial structure.

## 8.2 Analysis Methods

To determine the genetic structure of the biofilm I use the iDynoMiCS variables that relate to the genetic tree of the cells. Recalling from Chapter 3, these are the family number, the generation, and the genealogy. A unique family number is assigned as 1...n for each of the initial cells in the simulation. Cells divide by binary fission and so each initialised cell produces a genetically independent genetic tree such as that in Figure 8.1. The progeny of each of the initial cells inherit the same family number. The generation is defined as the number of divisions since the family was initialised, beginning with zero for the initial cells in the simulation. The final genetic variable, the genealogy number, is initialised as zero for the cells at the start of the simulation. Then, as a cell divides, one daughter cell retains the genealogy of the mother cell $G_{mother}$, while the other gets a new genealogy $G_{new}$, as $G_{new} = 2^{D_{new}-1} + G_{mother}$ where $D$ and $G$ are the generation and genealogy respectively. Taken together, these three parameters mean that each cell's position in the genetic tree is uniquely identified [83].

The simulations that I discuss in this section were produced using two different methods, which differ in the labelling of the genetic parameters. First, I performed an enhanced analysis of the same set of simulations which were discussed in previous chapters. In these simulations, only the family number was recorded as the genealogy and generation updates were not working, as discussed in Section 4.2 of Chapter 3. Later, and for most of this chapter, I discuss a new

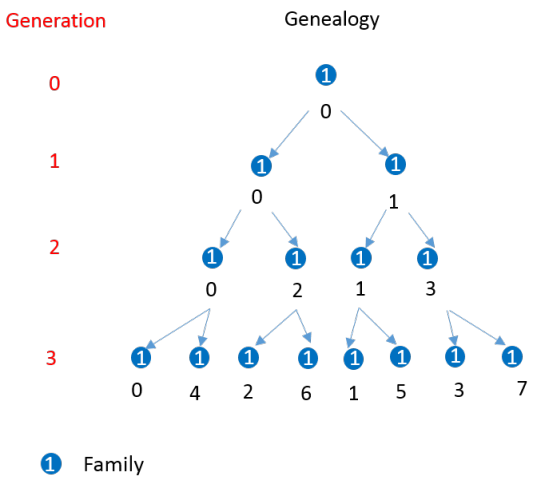



Figure 8.1: This Figure shows a sample genetic tree beginning from a single cell. The generations of the cells are shown on the left in red. The genealogy of each cell is shown in black below each cell. The family number of each cell is shown in white on top of the cell.

set of simulations which were produced using the clipping code in combination with the genetic number relabelling outlined in Section 4.3 of Chapter 3 so the full genetic information could be recalled for long time simulations. To recall, this involves relabelling the generation, genealogy and family at each simulation segment transition, i.e. at each 'clip', and also recording a conversion key. This procedure retains all the information necessary to identify each cell's position in the genetic tree, while decreasing the memory that is needed during the simulation and during later storage, which then allows me to perform long simulations retaining the full genetic information. In the rest of this section, I outline the methods used to analyse these simulations, beginning with an analysis which focuses on the family number and then moving on to analyses of the genealogy number.

### 8.2.1 Analysis of Family

The first calculations that I present here use the family number. While the family number alone cannot give us information about the entire genetic tree, since the family number does not identify each cell's position in the genetic tree uniquely, it is nevertheless a good starting point for investigating the genetic diversity of the biofilm. A particular family number labels the progeny of those cells that were present at the particular moment in time when the family number was assigned.

This could be when the simulation was initialised, or at the relabeling that occurs during a simulation segment in my clipping code. As the biofilm grows, some lineages are lost from the active layer and do not proliferate further. Therefore the number of distinct families in the active layer decreases in time. To analyse the dynamics of the genetic diversity, I track this loss of families, for biofilms with different morphologies. This is a good starting measure for seeing how the different phases (i.e. spatial structures) I identified in the previous chapters might impact the resulting genetic diversity of the biofilm.

### 8.2.2 Analysis of Genealogy

As we have seen above, the family number tells us which cells are the offspring of the cells that were present at some fixed time point - either the very start of the simulation, or at a the point of relabelling at the start of a simulation segment (depending which type of simulation I am talking about). However, the family number does not provide full information because it is not unique to each cell in the genealogical tree (Figure 8.1). To expand on the analysis above, I now also use the genealogy number in our analysis, which labels the unique position of each cell in the genetic tree.

The main calculation I focus on involving the genealogy is the genealogical distance between two cells. The genealogical distance is the number of divisions which separate two cells in the genetic tree. Figure 8.2 shows a sketch of what this looks like, where the genealogical distance between cells A and B can be calculated by counting the green arrows. This calculation involves the genealogy number and so is able to give us additional information compared to using the family number alone, but this comes at the expense of the calculations being rather more involved as well as more computationally intensive. In what follows, I firstly discuss the full calculation of the genealogical distance. I then discuss a modified version of this calculation which uses aspects of our relabelling technique (Section 4.3) to speed up the calculation.

#### Full calculation of the genealogical distance

As I have mentioned, the genealogical distance between a particular pair of cells is the number of divisions which separate the two cells in the genetic tree. For two cells in the same family where we know their genealogy and generation, their

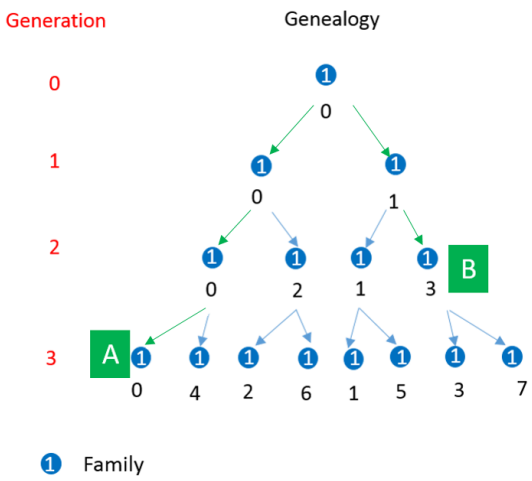



Figure 8.2: This sketch is to demonstrate the genealogical distance between two cells A and B. The genealogical distance is the number of divisions which separate two cells, or in other words the number of steps in the genealogical tree to get between two cells. In this example, between cells A and B there are five steps, which can be seen by courting the number of green arrows. The generations of the cells are shown on the left in red. The genealogy of each cell is shown in black below each cell. The family number of each cell is shown in white on top of the cell.

genealogical distance is calculated by summing the number of generations that separate each cell from their common ancestor as in Algorithm 4. This algorithm essentially reverses the calculations which are done to increment the genetic labels each time there is a division. Finding the generation of the previous cell in Step 1(a) simply involves subtracting one from the generation of the current cell. Finding the genealogy of the previous cell in Step 1(b) is a little more complicated. I recall from Figure 8.1, one of the daughters of a dividing cell takes the mother cell's genealogy number, while the other takes a new genealogy label as $G_{new} = 2^{D_{new}-1} + G_{mother}$ where $D$ and $G$ are the generation and genealogy respectively. Finding the previous genealogy number therefore involves working out whether the current cell took the genealogy label from its mother or got a new one. Looking again at Figure 8.1 we can see that the daughter cells which have the same genealogy as their mother cell can be identified because they have a genealogy which is less than half of the maximum possible genealogy number for that generation. The maximum genealogy number for a certain generation $D$ is $2^D$ (even if there are not enough cells to take up all these values), so this means we can infer what the mother cell's genealogy is from whether or not the genealogy is greater or less than half the maximum genealogy number which would be possible for that number of generations. Once this has been done, we see in Step 2 that I

look for the cell which is the common ancestor of the chosen two cells, and sum the number of generations which have passed between the common ancestor and each of the cells being compared.

---

**Algorithm 4** Genealogical distance between two cells $i = Cell1$ and $i = Cell2$ in the same family (i.e. $F^{1,0} = F^{2,0}$), with generation $D^{i,0}$ and genealogy $G^{i,0}$. The second subscript denotes the ancestry of the cell, e.g. the ancestors of cell 1 will have genealogy $G^{1,j}$ and generation $D^{1,j}$, and $j = 0$ denotes the initial cell being compared.

---

1. For cells $i = Cell1$ and $i = Cell2$, begin with the current cell and find the genetic information of all the cell's ancestors, working back one cell at a time.
   (a) Calculate generation of previous cell: $D_{i,j-1} = D_{i,j} - 1$
   (b) Calculate genealogy of previous cell:
      i. If $2^{D_{i,j}}/2 < G_{i,j}$; Then $G_{i,j-1} = D_{i,j} - 2^{G_{i,j}}$
      ii. Else If $2^{D_{i,j}}/2 \geq G_{i,j}$; Then $G_{i,j-1]} = G_{i,j}$
2. Count back the number of generations from each cell until the common ancestor of both the cells, and sum the number of generations that have passed for each cell.

---

For two cells which are in different families, I have to perform a slightly different calculation. The fact that cells are in different families means they are the progeny of different 'founder' cells at the start of the simulation, and so the two cells have no common ancestor within the simulation. This means I cannot calculate the genealogical distance in the same way as above. However, I can calculate the genealogical distance between each of the chosen cells and the initial cells of the simulation by performing Step 1 in Algorithm 4 and then summing the number of divisions to the initial cells in the simulation. This would be the genealogical distance if all cells at the start of the simulation were identical. Another way of saying this is that I am calculating the contribution to the genealogical distance which arises over the course of the biofilm growth, and not before. Results which use this assumption will be clearly indicated on the relevant plots used in the rest of the chapter.

In my long time simulations, I do not store the information for the entire trajectory directly, but instead I use the genetic information relabelling algorithm outlined in Chapter 4. This is because the genealogy number gets sufficiently large that it slows down the simulations. For calculating the genealogical distance,

this means that even though in principle I could undo the genetic information relabelling and then perform Algorithm 4, the genealogy numbers are sufficiently large that this is computationally difficult. Instead, I use a modified version of Algorithm 4 which makes use of the fact that the family numbers were relabelled at each simulation segment. I outline this in the next subsection and demonstrate the difference in speed of the calculations.

### Modified calculation of the genealogical distance

The modified genealogical distance calculation for simulations performed using the genetic information relabelling algorithm is outlined in Figure 8.3. It produces the same value as the full method, but in a way which is much faster. As with the simple calculation above, I compare the ancestors of a chosen pair of cells until I find their common ancestor, and work out how many generations each of the cells are away from this common ancestor. However, instead of completely unravelling all of the genetic parameters and using Algorithm 4, I use some the information which is gained during the relabelling process to avoid unravelling the genetic information, outlined in Figure 8.3.

In the example in Figure 8.3, cell A is the common ancestor of cell F (Cell 1) and cell G (Cell 2). We can see that in the simulation segment which contains the common ancestor (segment 0), the ancestors of cells 1 and 2 have the same (relabelled) family number, and they have a different family number in the other segments. We can then see that the genealogical distance between cell F and cell A can be calculated by adding up the number of generations which have occurred between cells F and D (2); D and B (2). The genealogical distance between cell G and A can be calculated in a similar manner. In this way, I do not use the genealogy to calculate the genealogical distance, except in Segment 0 where I calculate the distance between cells B and C using Algorithm 4, which is tractable since within one segment the genealogy numbers remain small. For this example, this means the total genealogical distance between cells 1 and 2 is 10 divisions. I note that this algorithm is non trivial to implement computationally because I have to correctly use the conversion keys to ensure I am comparing the correct cells with the correct number of relabels in the genetic tree. However, the implementation is worth the effort: for calculations in my simulated biofilms where there are many more divisions in each segment than in Figure 8.3, this speeds up the calculation significantly, as we shall see in the next subsection.

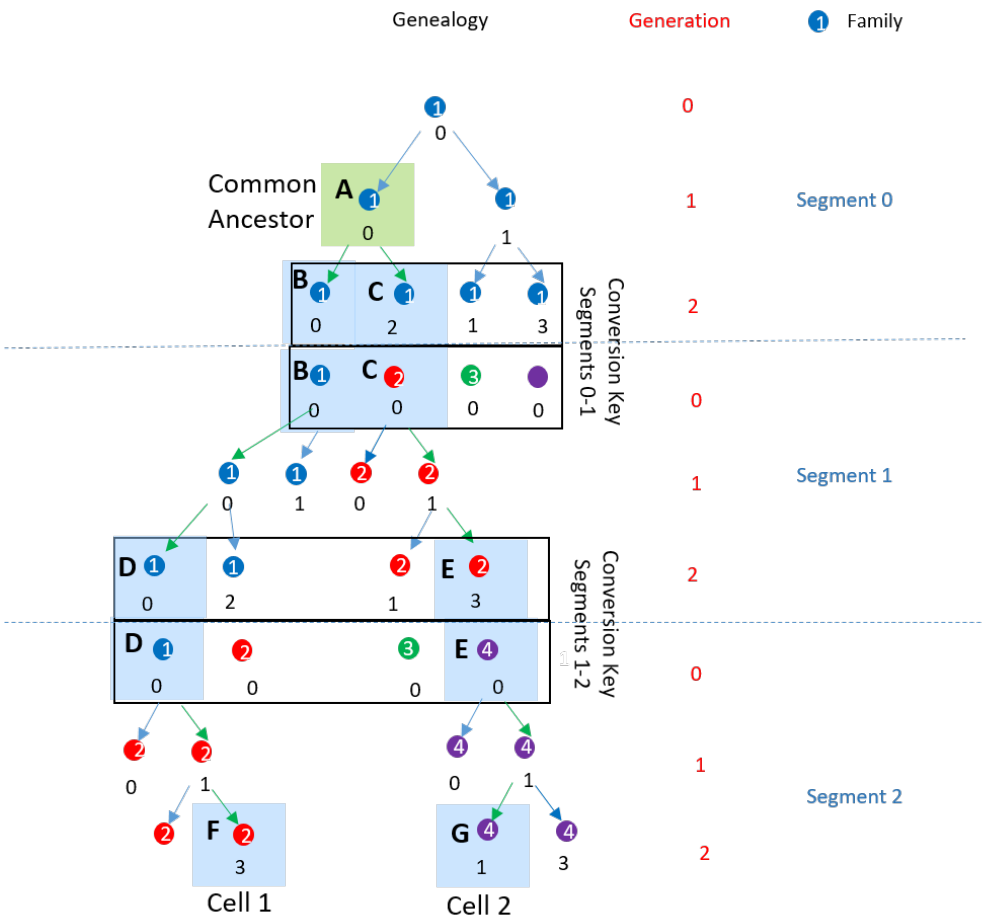



Figure 8.3: Diagram of how the modified genealogical distance calculation works, comparing cells F and G. Divisions to be counted are shown with green arrows. The common ancestor of cells F and G is cell A, highlighted with the green box. Family numbers local to the current segment are shown in white on top of the cell. Generation numbers local to the current segment are shown in red to the right hand side. Genealogy numbers are shown in black below each cell. In the black boxes are the conversion keys, which show how the cells are relabelled at the simulation segment transitions (for more detail see Chapter 4).

### Testing the Genealogical Distance Calculation

To confirm this modified algorithm works, I compare the genealogical distance between a given cell and 100 randomly selected cells in its surroundings at a number of different points in the trajectory of the biofilm for short times, calculated both with the full calculation described in Section 8.2.2 and with the modified calculation described in Section 8.2.2. All cells are selected randomly, and so this calculation is not intended to tell us anything other than whether or not the calculations produce the same results. In the two left hand panels of Figure 8.4, we can see that the average and standard deviations of the genealogical distance for these randomly selected pairs of cells produced using the two calculation methods (i.e the red and blue crosses are on top of each other.) In the right hand panel of Figure 8.4, we can see that the calculation times for the two calculation methods differ significantly. For larger biofilm sizes, each cell has significantly more ancestors whose genetic information must be compared in order to find the common ancestors. This means that in the modified calculation in Section 8.2.2, the ancestors only need to be compared explicitly for a single segment (the segment that contains the common ancestor) rather than explicitly for every ancestor, so the computational effort does not increase greatly with the length of the simulation.

So far I have presented the calculation for the genealogical distance between a chosen pair of cells. However, in order to compare the genealogical distance between cells over the spatial range of a biofilm and over the trajectory of biofilm growth, I will eventually need to average genealogical distances between many cells. In the interests of clarity, this is discussed later in Section 8.3.2.

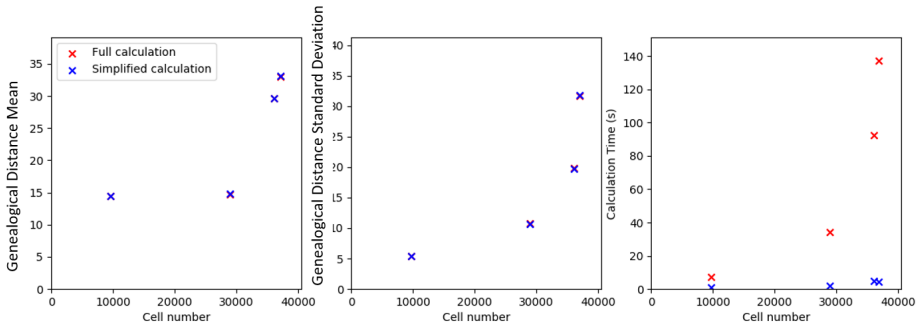



Figure 8.4: Comparing calculation methods for the genetic distance. The central and left panels show properties calculated using the two different genealogical distance methods, where we can see the values for the two calculations overlap as they produce the same values. In the the right hand panel, we can see the time taken to calculate the geneological distance between a cell and a number of neighbours is significantly different for the two different calculations, with the simplified calculation being significantly faster.

## 8.3 Results

### 8.3.1 Loss of initial cell lineages from the growing layer.

In this section, I present my results regarding the loss of initial cell lineages, tracked using the family number. As we have seen, each of the cells present at the start of the simulation are labelled with a unique family number, and the progeny of these inital cells retain the family number, allowing these cell lineages to be tracked. This means that I am able to track 300 separate lineages (corresponding to the 300 starting cells), and observe which of them remain in the growing layer of the biofilm over the course of biofilm growth. This approach of tracking the lineages of all starting cells contrasts with much of the previous work of this type, in which there are only two separate cell lineages which are tracked (eg [28, 35]).

I firstly examine simulation snapshots, coloured by family number, for different simulation times, as show in Figure 8.5. As I have discussed, a unique family number is assigned to each cell at the start of the simulation, and the progeny of these cells retain the initial cell's family number. The simulations used to make this Figure do not involve the genetic relabelling algorithm, and so the cells retain these family numbers throughout the simulation. In Figure 8.5, each family number is given a unique colour, and since each simulation begins with

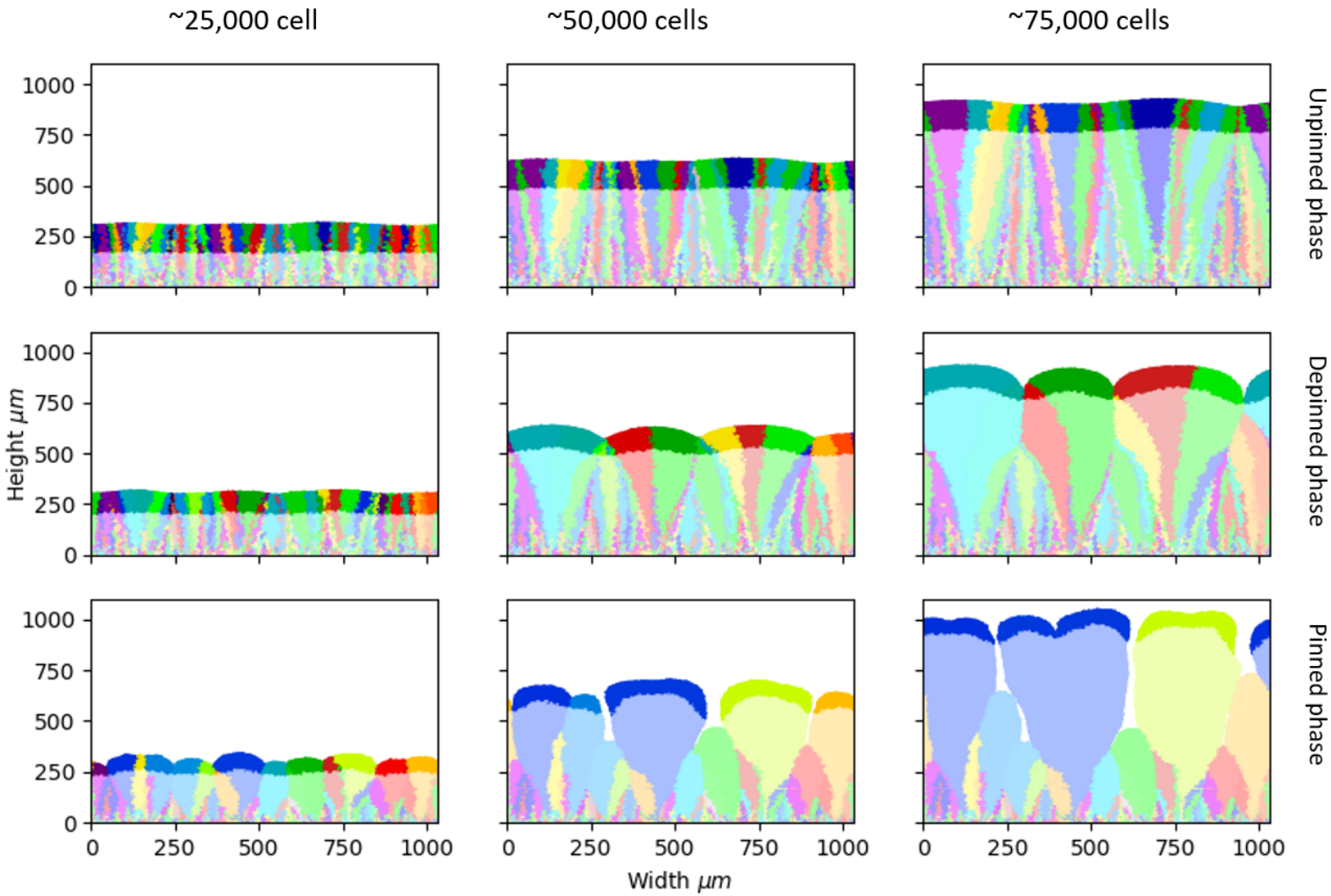



Figure 8.5: Family snapshot images. For the pin, depin and flat simulations from the main grid, at 25000, 50000 and 75000 cells each. Each colour represents the progeny of a different initial cell i.e. a different family number. The cells at the top of the biofilms which have a lower tranparency are cells which are in the active layer, as defined in Chapter 5.

300 cells, there are 300 colours present at the beginning. For cells which are no longer in the active layer in this plot, the colour has increased transparency (i.e. the darker colours at the top of the biofilm show the cells that are in the active layer). In all the simulations, we can see that some cell families come to dominate the growing front of the biofilm, while other families are left behind and remain only deep inside the (non-growing part of the) biofilm. This implies a loss of diversity within the active layer. It also appears that once a family has been lost from the active layer, it does not (or is at least very unlikely) to return to the active layer.

Comparing the different phases of biofilm grwoth in Figure 8.5, we can qualitatively see a number of differences between the phases. Though families are lost from the active layer in all the phases, it seems that families are lost from the active layer much more rapidly (in terms of generations) in the pinned phase, so that for the same size biofilm, the pinned phase has fewer families in the active layer. In the pinned phase we can clearly see that different families come to dominate each of the fingers, and so a given family becomes physically separated from other families. Looking carefully at the 'troughs' between biofilm fingers, we can see these pinning sites appear to be sites of diversity loss, as families are lost from the growing interface at this point. Interestingly, in the depinned phase, although families are not separated into different fingers, it appears that a gap or thin part of the active layer can have a similar affect of separating families from one another. This is most evident in the right hand snapshot of the depinned images, where we can see dividing lines between different families below thin parts of the active layer and no remixing of adjacent cells.

As an alternative way to view this diversity loss, in Figure 8.6 I combine the family information from multiple simulation times into one plot. For each time point, I indicate using colours the fraction of the population belonging to each family. Since each simulation begins with 300 cells each with a unique family number, there are 300 family colours which take up varying fractions of the overall biofilm. Putting different time points together along the horizontal axis gives us a picture of diversity loss in the biofilm as a function of time (or, as shown here, as a function of biofilm size). I have used biofilm size rather than time on the horizontal axis because biofilms in the pinned phase grow much slower than in the flat phase. We can see that a small number of families come to dominate the biofilm, as we expected from Figure 8.5. Comparing the plots in Figure 8.6 for the three phases, we see that diversity loss happens for smaller population sizes

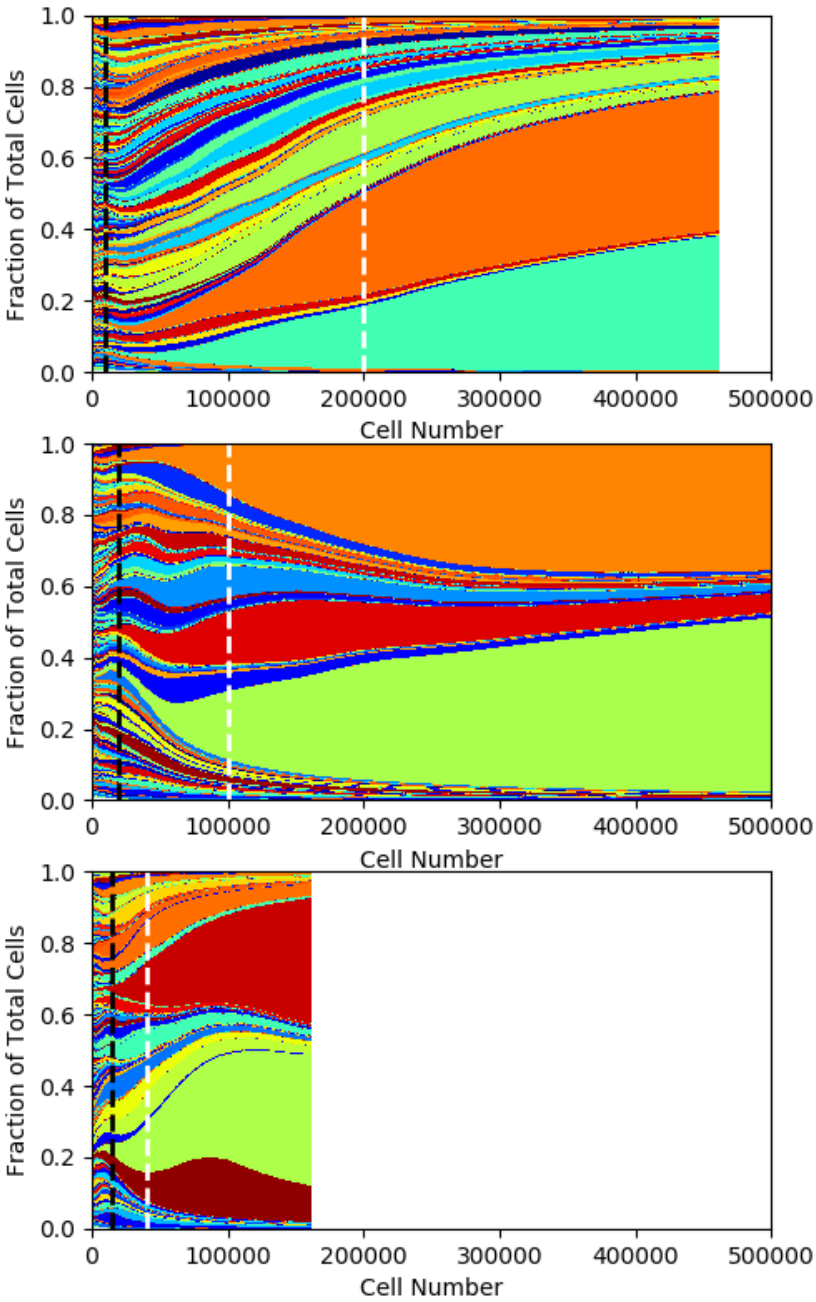



Figure 8.6: This Figure shows the fraction of cells in the biofilm which have each family number over the course of biofilm growth (i.e. the total cell number of the biofilm is on the x-axis and the fraction of those cells which have each of the family numbers is on the y-axis). Each colour represents a different family number. The black dashed line represents the cell number at which the biofilm becomes continuous across the simulation width, or in other words the average active layer thickness reaches it's steady state. The white dashed line represents the cell number when the standard deviation of the active layer thickness reaches its steady state.

in the rougher biofilms.

It is interesting to compare the amount of diversity loss in Figure 8.6 at the biofilm sizes where the average of the active layer thickness has reached a steady state, and where the standard deviation of the active layer thickness has reached a steady state. We recall from Chapter 5 that the average active layer thickness reaches a steady state early in biofilm growth, while the standard deviation reaches a steady state later in biofilm growth. The cell numbers at which these steady states are reached are shown with the black and white dashed lines in Figure 8.6 respectively. We can see that in all cases, a large number of families remain at the time at which the average active layer thickness reaches its steady state, and much of the diversity loss continues to happen after this. We can also see that even after the standard deviation of the active layer thickness has levelled off, several families continue to take up a larger and larger fraction of the biofilm. This is because the cell families which remain in the growing front can continue to expand their populations while families which are not in the growing front cannot, meaning they take up a smaller and smaller fraction of the overall population.

In Figure 8.6, I plotted the fraction of families in the entire biofilm. To focus on the families which remain in the growing front and so continue to expand their fraction of the biofilm, in Figure 8.7 I make a similar plot which focuses only on the families which remain on the interface. Figure 8.7 plots the family of each of the cells which are on the biofilm interface against their position on the interface (y-axis), as biofilm growth progresses (cell number on the x-axis). I note that for the pinned phase, where there interface has overhangs, there appear to be several families mixed in with each other. However, this is an artefact resulting from a 2D interface being collapsed onto one axis of this plot, so that cell families at the biofilm troughs appear alongside the cell families at the tops of the biofilm fingers. In all phases of biofilm growth, the decrease in time of the number of cell families and the taking over of the growing front by a just a few families is even more apparent in this plot (Figure 8.7) than in the previous plot (Figure 8.6). This implies that with time, a smaller number of cell families continue to have access to growing resources and hence with the ability to determine the future genetic fate of the biofilms. We can also see that the places where certain cell families fall behind the growing interface seems to be clustered at certain positions along the biofilm interface, which by comparison with Figure 8.5 I expect are due to gaps or thin parts of the active layer.

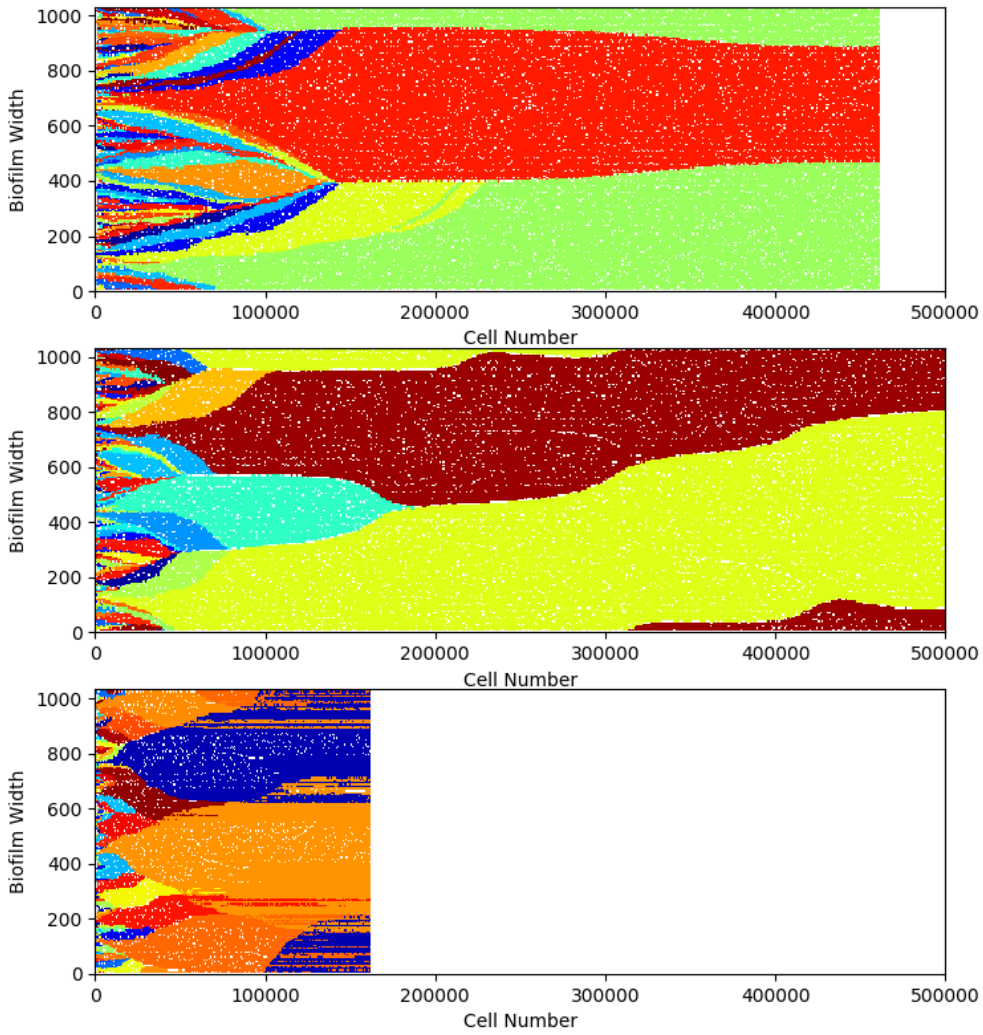



Figure 8.7: This figure shows the families which are present along the biofilm interface. The y-axis is the position along the biofilm width, and each colour represents the cell family which is present at that point. The x-axis is the cell number of the biofilm. The top panel is the unpinned phase, the central panel is the depinned phase and the bottom panel is the pinned phase.

Interestingly, in Figure 8.7 two cell families come to dominate the biofilms at large cell numbers for all three phases, though as we shall see later when I have introduced more complex methods for analysing the genetic diversity, this does not necessarily mean that the diversity of the resulting biofilms is the same. Though I begin with 300 cell family numbers corresponding with the 300 cells that are initialised, before I have reached any of the interesting steady state behaviour of the biofilm I am left with very few family numbers and this is not sufficient to understand if the unique behaviours of the different phases we have seen in the previous chapters are having an impact here. For example, in the depinned phase we can see a possible imprint of the active layer dynamics patterns we saw previously in Chapter 6 and in Figure 8.7. Specifically, the boundary between the two families at late times looks like part of the pattern of active layer gaps, but with only two cell families to distinguish between, any potential complexity of the pattern is lost. This motivates my decision in the next chapter to use the genealogy to find a measure of the ongoing diversity changes in the biofilm, not just the loss of initial diversity.

### 8.3.2 Genetic Diversity measured with Genealogical Distance.

In this section, I discuss the results involving genealogical distance between cells, or in other words the number of divisions which separate the two cells in the genealogical tree. This is a relevant measure of evolutionary distance because at each division event, there is the potential for a mutation, which means diversity can be gained at this point. This is particularly important in our analysis where I am examining biofilms of considerable size and the diversity which comes from cells dividing (and by implication accumulating mutations) cannot be ignored. The genealogical distance can also be measured between any two cells at any point in biofilm growth, and so this measure allows me to see how closely different parts of the biofilm are related to one another on an ongoing basis. This also means I can tell how cells within a single family are related.

(Recall calculation) I now present our results of the genealogical distance. We recall that there were two versions of this calculation depending on whether the two cells have a common ancestor within the simulation. Firstly, what I refer to as calculation 1 applies to the case where the two cells have a common ancestor and so the genealogical distance between the two cells can be calculated. Secondly, what we refer to as calculation 2 applies when the two cells originate from a different starting cell, and so we can calculate the genealogical distance only as far back as those starting cells, which we can think of the genetic diversity gained since the start of the simulation. Genealogical distance provides an indication of the genetic diversity which is gained by cells during the course of biofilm growth, as each division event implies a possible mutation.

In the same manner as we did for the family analysis, I begin with snapshot images to get a qualitative sense of what is going on for the different phases. In Figure 8.8 we can see the genealogical distance between a focal cell (coloured in red) and all the other cells. The other cells are coloured according to their genealogical distance from the focal (red) cell. The green colours represent cells with a common ancestor with the focal cell (calculation 1) and the blue colours represent cells without a common ancestor with the focal cell (calculation 2). The central cell has been selected at the minima of the interface in the left hand panel, and at the maxima on the right hand panels. The top, central and bottom panels correspond to the upinned, depinned and pinned phases respectively, each with configurations of approximately 75000 cells. These are just $200\mu m$ portions

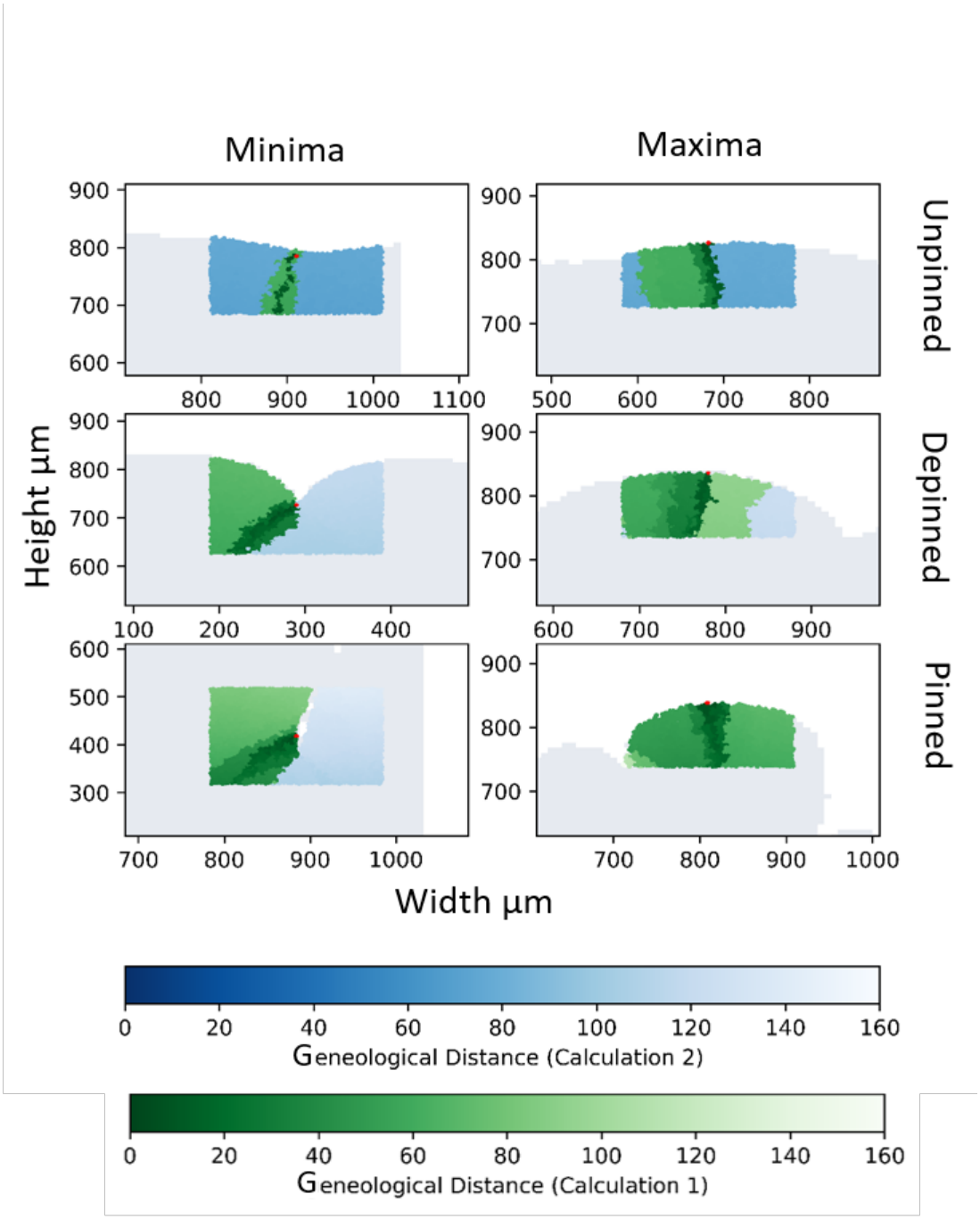



Figure 8.8: This Figure shows the genealogical distance between the central cell, shown in red, and the surrounding cells for parts of the 75000 cell biofilm for each of the three phases. The cells in green represent the cells which have a common ancestor in the simulation and so the genealogical distance is calculated using Calculation 1, as explained in the text. The blue cells represent cells which do not have a common ancestor in the simulation, and so the genealogical distance is calculated using Calculation 2. The left panels show part of the biofilm at its maximum interface point, and the right hand panels show part of the biofilm at its minimum point.

of the overall biofilm configurations, because even with the genealogical distance calculation optimised as we saw in Section 8.2, it is still a hefty calculation.

In Figure 8.8 it is possible to see the differences between the results for focal cells at the maxima and minima of the biofilm interface. In the depinned and pinned phases, we can see that cell lineages are lost at the interface minima by the shades of green that are lost there. This conforms what we saw in Figure 8.5 - that gaps in the active layer are sites of diversity loss. In the unpinned phase, we can see that even small variations on the biofilm interface limits access to the growing front, and so some diversity loss. At the maxima, we can see that the lineages that are expanding, and so increasing their diversity. This implies that the overall diversity of the biofilm is a trade off between gains in diversity at the growing interface and especially the maxima, and losses in diversity which occur at the minima of the growing interface and at gaps in the active layer. It is worth reiterating that the process via which diversity is gained (cell division and mutation) is different to the process via which diversity is lost (loss of lineages behind the growing front), and the resulting diversity trade off will depend both on the mutation rate and all the processes which influence the active layer gap behaviour which I have highlighted in the previous chapters.

I now move to looking more carefully at the differences between the phases by examining the genealogical distance along the interface in Figure 8.9. Here we can see biofilm snapshots at 50000 cells for each of the phases, with the genealogical distance from the focal cell to each of the cells along the active interface being represented by a colour-map. The genealogical distance is also represented in the scatter plot above each of these snapshots, with each point representing the average genealogical distance from the focal cell for cells in $20\mu m$ sections of the interface. The cells which have a common ancestor with the central (red) cell are represented in green, while the cells that do not are in blue.

There are two key features of Figure 8.9, one for cells which have their common ancestor in the simulation, and one for cells which do not. Firstly, the maximum genealogical distance in each case occurs for the cells which do not have a common ancestor with the central cell i.e. the blue 'wings' in the top panels. Here we can see that for a similar size biofilm, the cells on the interface have undergone many more divisions for the rougher biofilms than for smooth ones. This means that cells at the growing interface of the pinned phase could be more diverse than those at the growing interface of the unpinned phase, despite being separated from other cell lineages in other biofilm fingers. Secondly, we look at the distribution

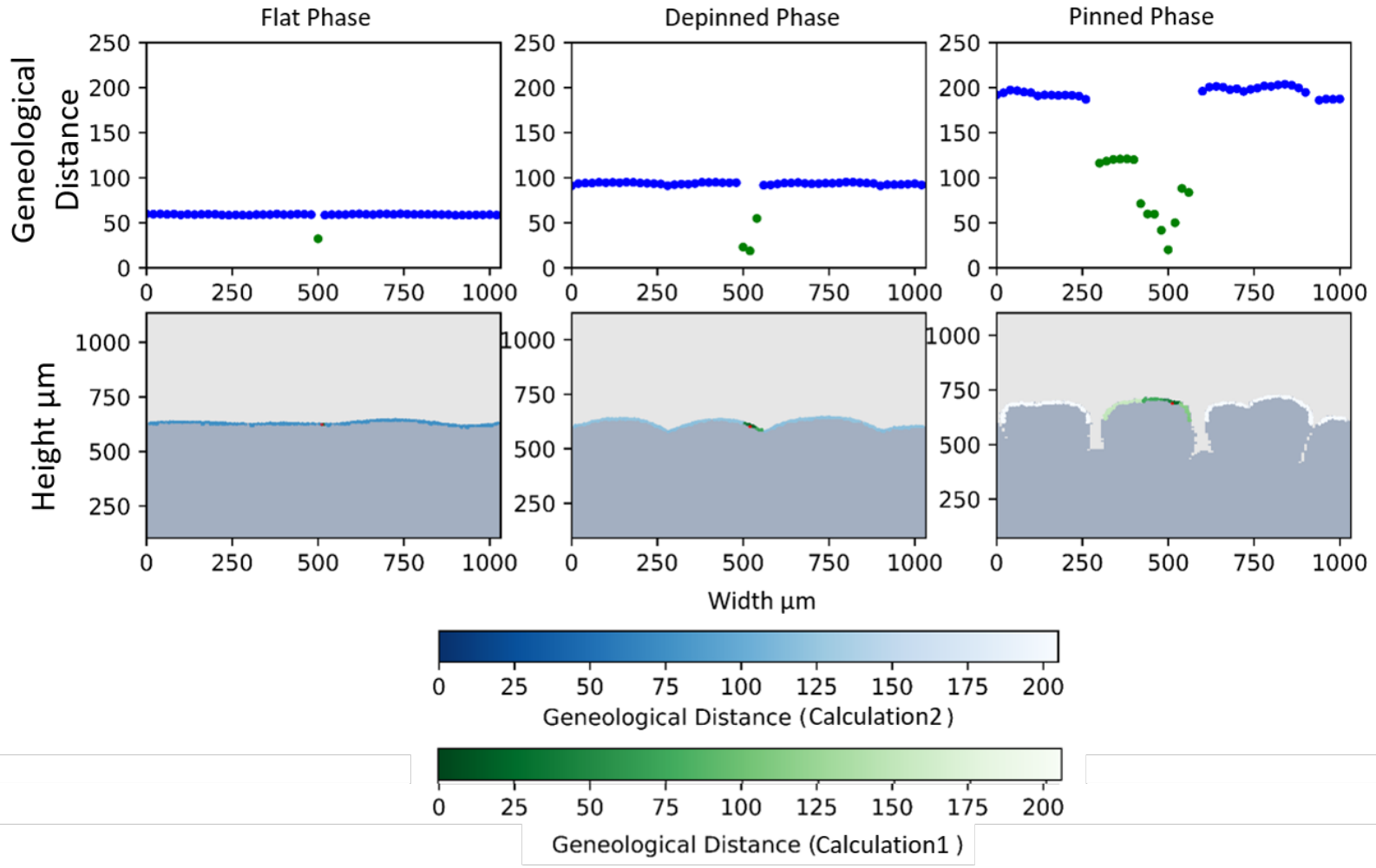





Figure 8.9: The bottom panels of this figure is showing the genealogical distance between the central cell and all the other cells on the active interface. The cells in green represent the cells which have a common ancestor in the simulation and so the genealogical distance is calculated using Calculation 1, as explained in the text. The blue cells represent cells which do not have a common ancestor in the simulation, and so the genealogical distance is calculated using Calculation 2. The top panels of this plot shows averages of cells in $20\mu m$ increments of those cells plotted in the snapshot below them.

of genealogical distances between cells with a common ancestor in the simulation i.e the green'V' shape part of the genealogical distance. The broader 'V' shapes in the pinned phase implies a different pattern of relatedness among cells at the interface in the unpinned compared to the pinned phase. Specifically, the genealogical distances between the cells are more uniformly distributed in the unpinned phase.

Finally, I average the maximum genealogical distance over the course of the simulations in Figure 8.10. This means that I am considering the behaviour of the cells along the interface without a common ancestor (i.e. blue 'wings' of Figure 8.10) over the course of biofilm growth. For each time point in the biofilm growth, I perform three repeats of selecting a random cell along the active interface and calculating the genealogical distance between it and the other cells along the active interface, before averaging. Therefore I obtain a trajectory of average relatedness along the biofilm interface, for each simulation. This confirms what we saw for a single time point in Figure 8.9, that along the interface of the pinned phase, the cells have undergone more divisions on average, meaning their potential number of mutations is higher and so these cells will be more diverse than cells along the interface of the other phases. Together with what we saw in the previous Section 8.3.1 this highlights that the overall diversity of the biofilm will be a trade off between the diversity which is present as a result of the starting cells and diversity which is built up over the course of biofilm growth.

## 8.4 Discussion

In this chapter I have investigated the genetic diversity of the unpinned, depinned and pinned phases which I defined in the previous chapters. We saw that that in all phases, the active layer acts as a population bottleneck, as only cells within the active layer can grow. This is consistent with several other studies, where it is argued that this sampling effect on the population increases the chance that a neutral of deleterious mutation will come to dominance within the biofilm [75, 77]. We also saw that different points across the biofilm had different roles in terms of population diversity. Specifically, we firstly saw that pining sites are sites of diversity loss, where many cell lineages can be lost in one go. We secondly saw the peaks of biofilms are places where the most diversity is gained due cell divisions (and therefore potential mutations) and that the cell divisions were most concentrated at the maxima for the pinned phase. In summary, the

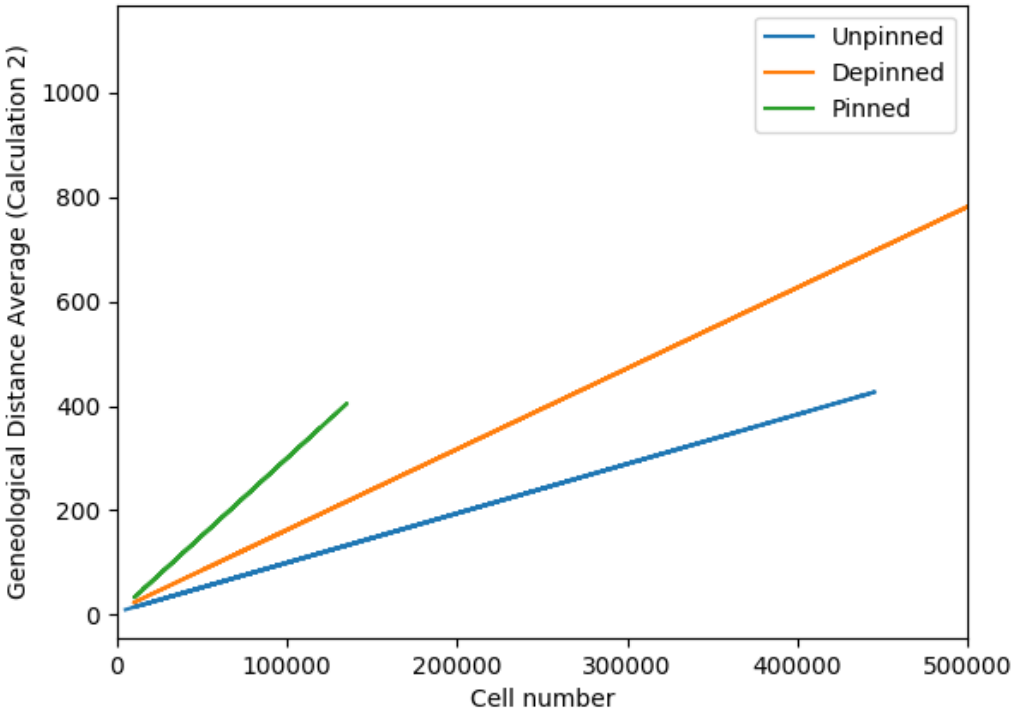



Figure 8.10: This Figure shows the average genealogical distance between cells on the interface with no common ancestor within the simulation i.e. descended from different initial starting cells, calculated using calculation 2. This corresponds to averaging the blue elements of Figure 8.9.

diversity across different points in the biofilm as well as the overall diversity will depend on the trade off between these various processes.

An interesting question which arises out of my work is whether there is any difference between a gap in the active layer and a pinning site in terms of the segregation of different cell lineages. Although it is intuitive to think about the genetic diversity of rougher biofilms being lower because different cell lineages become separated into different biofilm fingers, my work does not conclusively show if the separation of different cell lineages into different biofilm fingers (resulting from a pinning site) would have a significantly different impact than cells being separated due to a gap in the active layer. We saw that when cell lineages (or family numbers) fall out of the active layer, they are not able to remix with the active layer or with each other, and we saw that gaps in the active layer also cause fixed boundaries between different cell lineages. However, one difference is that in the depinned phase, the gaps in the active layer move around a great deal (as we saw in Chapter 6), meaning that the cell lineages are not consistently lost from one area of the biofilm as they are in the pinned phase. This could have a significant impact, which I was not able to elucidate with the measures of diversity which I investigated here. It is also the case that if the conditions of biofilm growth changed, for example more nutrients became available or flow became a significant factor, remixing of cell lineages would be significantly easier for biofilms where lineages were not not separated into different fingers (i.e resulting from a gap in the active layer rather than a pinning site). I think it would be interesting to investigate further the evolutionary differences between active layer gaps and pinning sites.

An important further question is whether there are clear differences between active layer gaps and pinning sites is whether the differences between phases in genetic spatial structure. While we saw in our previous work clear phase behaviour in the macroscopic spatial structure, whether there are qualitatively distinct genetic behaviours occurring in the different phases here is not conclusive from our work, though there is some evidence. We saw in Chapter 2 that Nadell *et. al.* argue that the segregation of two different cell lineages varies smoothly with the active layer thickness across a range of biofilm shapes. It is possible that this is the case in my work, though I think it is more likely that there are distinct phases in the genetic spatial structure that the work I have done here is not able to elucidate. We have seen some evidence that there are different behaviours at peaks and troughs and that movements of the active layer gaps matter. I

also think that the sensitivity to the local thickness of the active layer of cell lineages dropping out of the active layer and so having their expansion paths blocks means that it is possible that there is a sensitivity to the active layer thickness fluctuations in a similar manner to the morphological phases described in the previous chapters.

Assuming the mutation rate is such that divisions mean a significant accumulation in genetic diversity in the particular case we are looking at, the differences in diversity accumulated by cell division and mutation have implications for what happens when selective pressures are applied. For example, Pamp *et. al.* show that antibiotic treatment of biofilms *Pseudomonas aeruginosa* can kill either the growing cells on the tops of the biofilms or the non-growing cells on the inside of the biofilm, depending on the antibiotic. For the pinned phase in my simulations where there is much greater diversity at the peaks than the troughs or inside the biofilm, antibiotics primarily killing the outside or the inside would affect the genetic make up on the bacteria which remain. If the antibiotics were to kill the growing portions of the biofilm, this would leave behind a much less diverse population than if the inactive cells inside the biofilm were killed. The same would be true for the unpinned phase, though as the distribution of divisions and so diversity accumulated by mutation is more uniform, this is likely to be a smaller effect.

## 8.5 Summary

In this chapter I have investigated the genetic spatial structure for the three macroscopic spatial structure phases I established in the previous chapters. I found that cell lineages drop out of the active layer as biofilm growth progresses, and so genetic diversity which was present at the start of biofilm growth is lost. I also observed that diversity is gained through cell division and the corresponding accumulation of mutations in the active layer, and so the diversity of the biofilm is a trade off between these processes. I found that pinning sites are key sites of diversity loss as this is where many cell lineages are blocked from further expansion, while the tops of biofilm fingers are sites of diversity expansion as this is where cell divisions are most concentrated. I argued that further work would be needed to establish if there was clear transitions in the diversity behaviour in a manner that corresponds to the macroscopic spatial structure.

# Chapter 9

# Conclusion

I believe that the key contribution of my thesis has been to produce a body of work which investigates both macroscopic spatial structure (i.e the biofilm roughness) and the biofilm genetic spatial structure (i.e the fates of various cell lineages). I recall from Chapter 1 that an important question in biofilm research is how biofilm structure is linked to biofilm behaviour or other aspects of biofilm growth. This has motivated a great number of studies attempting to produce a phase diagram of biofilm morphology, in part as a means to unpick the feedback loop between biofilm structure and function. The genetic spatial structure is one such example of this feedback loop between morphology and function, though there have been relatively few studies seeking to understand this relationship. In my work, I was able to establish three phases of biofilm growth, and then to study the behaviour of the genetic spatial structure in each of these phases. While there remains work to be done in terms of quantifying the difference in genetic spatial structure between the different phases, I was able to demonstrate the necessity of studies which more clearly link biofilm morphology to biofilm genetic spatial structure.

My approach to linking the macroscopic spatial structure to the genetic spatial structure has been to focus on the active layer. As we have seen, the active layer is the actively growing cells usually at the top of the biofilm. I saw three phases in the biofilm phase behaviour that had different behaviours of the active layer and the pinning behaviour that results. Finally, I observed a pinned phase in which pinned sites appear, or in other words the active layer becomes discontinuous so parts of the interface become stationary and lag behind the growing front.

Secondly, I observed a depinned phase which has pinning events which from but can ultimately be overcome. Finally, I observed an unpinned phase which has a continuous active layer and no pinning sites. When turning to link this to the genetic spatial structure, I again used the active layer. The cell lineages which are in the active layer are the ones which come to dominate the biofilm as they have access to nutrients and are able to grow and divide. This means the active layer is also where diversity is gained due to mutations that occur during cell division. I found that pinning sites are key sites of diversity loss as these are where cell lineages drop out of the active layer and are blocked from any further expansion. I think that while there have been studies which focus on the active layer as a driver of macroscopic spatial structure and as an important genetic bottleneck, these phenomena have rarely been studied in tandem and I believe that my work has demonstrated the effectiveness of this active layer-based approach to studying biofilm spatial structure.

While I am confident in the qualitatively distinct nature of the phases I observe in my simulations, I think that quantitative classification of these phases requires further work. I argued in Chapters 6 and 7 that while it may in principle be possible to classify these phases using scaling exponents and an interface growth theory approach, searching for an order parameter is more conducive for understanding the mechanisms driving the roughness and therefore to linking biofilm morphology to genetic spatial structure. In my work I showed both qualitatively and quantitatively that the active layer dynamics as well as the average active layer alone matter in the resulting phase behaviour. I argued that the mechanical interactions play an important role in the active layer dynamics, as the actively growing cells are the cells which are interacting mechanically the most, and these interactions 'push' around the gaps in the active layer and play a role in closing up pinning sites in the depinned phase. I observed in Chapter 5 that a number of input parameters related to the nutrient dynamics had distinctive and quantifiable impact on the active layer, but I was not able to produce a combined parameter to predict the average active layer depth in a similar manner to Nadell *et. al.* [28]. I believe that producing similar plots for simulation input parameters which define the mechanical interactions and combining these with the parameters which define the nutrient behaviour would be a productive way forward to finding a control parameter with predictive power. I think this would also be an important contribution to a literature which is increasingly recognising the role of mechanical interactions in biofilm behaviour [14, 16].

I believe the main limitations of my work have come from the large scale of the simulations required to reach the steady state of biofilm growth. While improvements in computationally efficiency over recent decades have made possible individual-based simulations in many fields of research to study emergent behaviour, there is still work to be done on computational efficiency (both in the software and in research practices). The iDynoMiCS simulation software is also not designed for long time simulations, it is written in Java and outputs files in XML format, and so despite the strenuous efforts to speed up the calculations, they were still rather unwieldy. Specifically in my work, this meant that I was not able to perform simulations over a particularly large parameter space for the phase diagram e.g. I did not perform any long time simulations changing the diffusivity or height of the diffusion boundary layer, which my early work had shown to have a clear impact on the active layer thickness. I was also not able to perform repeat simulations with different simulation sizes to determine any role for finite size effects. It also meant that many of the other ideas for expanding and exploring the work here over the course of my PhD remained on the drawing board, or were begun only to find they were computationally infeasible. Overall, I think this significantly reduced the work I was able to produce. Options for future work along the lines of my research may be to change to a different simulation software. A good option could be the biofilm simulation software recently released by Jayathilake *et. al.*, which uses a microbiological adaptation of the open source Large-scale Atomic/Molecular Massively Parallel Simulator (LAMPS) to significantly increase the computational tractability of large biofilm simulations [105]. Some authors have used an unrealistic value for the diffusion in their simulations to reach long times, and this avenue to achieving faster simulations could also be further explored.

In conclusion, in my thesis I have investigated both the macroscopic spatial structure and genetic spatial structure and have attempted to link them through the detailed study of the active layer and the production of a phase diagram. I was able to observe three distinct phases - and unpinned phase, a depinned phase and a pinned phase and to detail the distinctive interface roughness behaviours and active layer behaviours dynamics in each of these phases. I was able to qualitatively link the the genetic spatial structure to the pinning behaviour I observed in each of the phases and find an important role of pinning sites in terms of diversity loss, I did not have time to find an appropriate way of averaging my measures of genetic diversity and so to quantitatively link them to the morphology phase behaviour I observed. I also provided detailed discussions

of the contribution of my work to current debates in the literature.

# Bibliography


[1] James A. Shapiro. The significances of bacterial colony patterns. *BioEssays*, 17(7):597–607, 1995.

[2] P Watnick and R Kolter. Biofilm, city of microbes. *Journal of bacteriology*, 182(10):2675–9, 2000.

[3] Philip S. Stewart and Michael J. Franklin. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology*, 6(3):199–210, 2008.

[4] Apollo Stacy, Luke McNally, Sophie E. Darch, Sam P. Brown, and Marvin Whiteley. The biogeography of polymicrobial infection. *Nature Reviews Microbiology*, 14(2):93–105, 2015.

[5] Luanne Hall-Stoodley, J. William Costerton, and Paul Stoodley. Bacterial biofilms: from the Natural environment to infectious diseases. *Nature Reviews Microbiology*, 2(2):95–108, 2004.

[6] Shama Sehar and Iffat Naz. Role of the Biofilms in Wastewater Treatment. In *Microbial Biofilms - Importance and Applications*. InTech, 2016.

[7] M. Batté, B.M.R. Appenzeller, D. Grandjean, S. Fass, V. Gauthier, F. Jorand, L. Mathieu, M. Boualam, S. Saby, and J.C. Block. Biofilms in Drinking Water Distribution Systems. *Reviews in Environmental Science and Bio/Technology*, 2(2-4):147–168, 2003.

[8] H. Trøstrup, T. Bjarnsholt, K. Kirketerp-Møller, N. Høiby, C. Moser, N. HoiIby, and C. Moser. What Is New in the Understanding of Non Healing Wounds Epidemiology, Pathophysiology, and Therapies. *Ulcers*, 2013:1–8, 2013.

[9] R. O. Darouiche. Device-Associated Infections: A Macroproblem that Starts with Microadherence. *Clinical Infectious Diseases*, 33(9):1567–1572, 2001.

[10] C.A. Fux, J.W. Costerton, P.S. Stewart, and P. Stoodley. Survival strategies of infectious biofilms. *Trends in Microbiology*, 13(1), 2005.

[11] Don Monroe. Looking for Chinks in the Armor of Bacterial Biofilms. *PLoS Biology*, 5(11):e307, 2007.

[12] Eshel Ben-Jacob, Inon Cohen, Ido Golding, David L. Gutnick, Marianna Tcherpakov, Dirk Helbing, and Ilan G. Ron. Bacterial cooperative organization under antibiotic stress. *Physica A: Statistical Mechanics and its Applications*, 282(1):247–282, 2000.

[13] Oana Ciofu, Bente Riis, Tacjana Pressler, Henrik Enghusen Poulsen, and Niels Høiby. Occurrence of hypermutable Pseudomonas aeruginosa in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. *Antimicrobial agents and chemotherapy*, 49(6):2276–82, 2005.

[14] Dmitri Volfson, Scott Cookson, Jeff Hasty, and Lev S Tsimring. Biomechanical ordering of dense cell populations. *Proceedings of the National Academy of Sciences of the United States of America*, 105(40):15346–51, 2008.

[15] Rosalind Allen and Bartlomiej Waclaw. Antibiotic resistance: a physicist's view. *Physical Biology*, 13(4):045001, 2016.

[16] F. D. C. Farrell, O. Hallatschek, D. Marenduzzo, and B. Waclaw. Mechanically Driven Growth of Quasi-Two-Dimensional Microbial Colonies. *Physical Review Letters*, 111(16):168101, 2013.

[17] Miguel Trejo, Carine Douarche, Virginie Bailleux, Christophe Poulard, Sandrine Mariot, Christophe Regeard, and Eric Raspaud. Elasticity and wrinkled morphology of Bacillus subtilis pellicles. *Proceedings of the National Academy of Sciences of the United States of America*, 110(6):2011–6, 2013.

[18] Eshel Ben-Jacob, Inon Cohen, Ofer Shochet, Adam Tenenbaum, András Czirók, and Tamás Vicsek. Cooperative Formation of Chiral Patterns during Growth of Bacterial Colonies. *Physical Review Letters*, 75(15):2899–2902, 1995.

[19] Cristian Picioreanu, Jan-Ulrich Kreft, and Mark C M Van Loosdrecht. Particle-based multidimensional multispecies biofilm model. *Applied and environmental microbiology*, 70(5):3024–3040, 2004.

[20] J. W. T. Wimpenny. The Growth and Form of Bacterial Colonies. *Journal of General Microbiology*, 114(2):483–486, 1979.

[21] Juan A. Bonachela, Carey D. Nadell, João B. Xavier, and Simon A. Levin. Universality in Bacterial Colonies. *Journal of Statistical Physics*, 144(2):303–315, 2011.

[22] Ana Margarida Sousa, Idalina Machado, Ana Nicolau, and Maria Olívia Pereira. Improvements on colony morphology identification towards bacterial profiling. *Journal of Microbiological Methods*, 95(3):327–335, 2013.

[23] C Picioreanu, M C van Loosdrecht, and J J Heijnen. A theoretical study on the effect of surface roughness on mass transport and transformation in biofilms. *Biotechnology and bioengineering*, 68(4):355–69, 2000.

[24] Carey D. Nadell, Knut Drescher, and Kevin R. Foster. Spatial structure, cooperation and competition in biofilms. *Nature Reviews Microbiology*, 14(9):589–600, 2016.

[25] Yun Shen, Guillermo L. Monroy, Nicolas Derlon, Dao Janjaroen, Conghui Huang, Eberhard Morgenroth, Stephen A. Boppart, Nicholas J. Ashbolt, Wen-Tso Liu, and Thanh H. Nguyen. Role of Biofilm Roughness and Hydrodynamic Conditions in Legionella pneumophila Adhesion to and Detachment from Simulated Drinking Water Biofilms. *Environmental Science & Technology*, 49(7):4274–4282, 2015.

[26] Isabel Frost, William PJ Smith, Sara Mitri, Alvaro San Millan, Yohan Davit, James M Osborne, Joe M Pitt-Francis, R Craig MacLean, and Kevin R Foster. Cooperation, competition and antibiotic resistance in bacterial colonies. *The ISME journal*, 12(6):1582–1593, 2018.

[27] Anton Kan, Ilenne Del Valle, Tim Rudge, Fernán Federici, and Jim Haseloff. Intercellular adhesion promotes clonal mixing in growing bacterial populations. *Journal of The Royal Society Interface*, 15(146):20180406, 2018.

[28] Carey D. Nadell, Kevin R. Foster, João B. Xavier, WT Vetterling, and AS Griffin. Emergence of Spatial Structure in Cell Groups and the Evolution of Cooperation. *PLoS Computational Biology*, 6(3):e1000716, 2010.

[29] Daniel S. Esser, Johan H. J. Leveau, and Katrin M. Meyer. Modeling microbial growth and dynamics. *Applied Microbiology and Biotechnology*, 99(21):8831–8846, 2015.

[30] Tolker-Nielsen and Molin. Spatial Organization of Microbial Biofilm Communities. *Microbial ecology*, 40(2):75–84, 2000.

[31] Pushpita Ghosh, Jagannath Mondal, Eshel Ben-Jacob, and Herbert Levine. Mechanically-driven phase separation in a growing bacterial colony. *Proceedings of the National Academy of Sciences of the United States of America*, 112(17):E2166–73, 2015.

[32] D. A. Head. Linear surface roughness growth and flow smoothening in a three-dimensional biofilm model. *Physical Review E*, 88(3):032702, 2012.

[33] Albert-Laaszlo. Barabasi and H. Eugene Stanley. *Fractal concepts in surface growth*. Press Syndicate of the University of Cambridge, 1995.

[34] Timothy Halpin-Healy and Yi-Cheng Zhang. Kinetic roughening phenomena, stochastic growth, directed polymers and all that. aspects of

multidisciplinary statistical mechanics. *Physics reports*, 254(4-6):215–414, 1995.

[35] Oskar Hallatschek, Pascal Hersen, Sharad Ramanathan, and David R. Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, 2007.

[36] Oskar Hallatschek and David R Nelson. Life at the front of an expanding population. *Evolution: International Journal of Organic Evolution*, 64(1):193–206, 2010.

[37] M. Matsushita, J. Wakita, H. Itoh, I. Ràfols, T. Matsuyama, H. Sakaguchi, and M. Mimura. Interface growth and pattern formation in bacterial colonies. *Physica A: Statistical Mechanics and its Applications*, 249(1):517–524, 1998.

[38] Jun-ichi Wakita, Kenji Komatsu, Akio Nakahara, Tohey Matsuyama, and Mitsugu Matsushita. Experimental Investigation on the Validity of Population Dynamics Approach to Bacterial Colony Formation. *Journal of the Physical Society of Japan*, 63(3):1205–1211, 1994.

[39] Rie Tokita, Takaki Katoh, Yusuke Maeda, Jun-ichi Wakita, Masaki Sano, Tohey Matsuyama, and Mitsugu Matsushita. Pattern Formation of Bacterial Colonies by E. Coli. *Journal of the Physical Society of Japan*, 78(7):074005, 2009.

[40] Akio Nakahara, Yuji Shimada, Jun-ichi Wakita, Mitsugu Matsushita, and Tohey Matsuyama. Morphological Diversity of the Colony Produced by Bacteria Proteus Mirabilus. *Journal of the Physical Society of Japan*, 65(8):2700–2706, 1996.

[41] Inon Cohen, Ilan G Ron, and Eshel Ben-Jacob. From branching to nebula patterning during colonial development of the Paenibacillus alvei bacteria. *Physica A: Statistical Mechanics and its Applications*, 286(1):321–336, 2000.

[42] M Matsushita, J Wakita, H Itoh, K Watanabe, T Arai, T Matsuyama, H Sakaguchi, and M Mimura. Formation of colony patterns by a bacterial cell population. *Physica A: Statistical Mechanics and its Applications*, 274(1):190–199, 1999.

[43] Shanika A Crusz, Roman Popat, Morten Theil Rybtke, Miguel Cámara, Michael Givskov, Tim Tolker-Nielsen, Stephen P Diggle, and Paul Williams. Bursting the bubble on bacterial biofilms: a flow cell methodology. *Biofouling*, 28(8):835–842, 2012.

[44] Azadeh Ghanbari, Jaber Dehghany, Timo Schwebs, Mathias Müsken, Susanne Häussler, and Michael Meyer-Hermann. Inoculation density and nutrient level determine the formation of mushroom-shaped structures in Pseudomonas aeruginosa biofilms. *Scientific Reports*, 6(1):32097, 2016.

[45] F Family and T Vicsek. Scaling of the active zone in the Eden process on percolation networks and the ballistic deposition model. *Journal of Physics A: Mathematical and General*, 18(2):L75–L81, 1985.

[46] Benoit B Mandelbrot. Self-Affine Fractals and Fractal Dimension. *Physica Scripta*, 32(4):257–260, 1985.

[47] Simon A Levin, JA Bonachela, and CD Nadell. Mathematical and computational challenges in the study of complex adaptive microbial systems. In *The Social Biology of Microbial Communities: Workshop Summary*. National Academies Press, 2012.

[48] Jun-ichi Wakita, Hiroto Itoh, Tohey Matsuyama, and Mitsugu Matsushita. Self-Affinity for the Growing Interface of Bacterial Colonies. *Journal of the Physical Society of Japan*, 66(1):67–72, 1997.

[49] Kazumasa A Takeuchi. An appetizer to modern developments on the kardar–parisi–zhang universality class. *Physica A: Statistical Mechanics and its Applications*, 504:77–105, 2018.

[50] Timothy Halpin-Healy and Yi-Cheng Zhang. Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics. *Physics Reports*, 254(4-6):215–414, 1995.

[51] Timothy Halpin-Healy and Kazumasa A Takeuchi. A kpz cocktail-shaken, not stirred... *Journal of Statistical Physics*, 160(4):794–814, 2015.

[52] Silvia N Santalla, Javier Rodríguez-Laguna, José P Abad, Irma Marín, María del Mar Espinosa, Javier Muñoz-García, Luis Vázquez, and Rodolfo Cuerno. Nonuniversality of front fluctuations for compact colonies of nonmotile bacteria. *Physical Review E*, 98(1):012407, 2018.

[53] C. Giverso, M. Verani, and P. Ciarletta. Branching instability in expanding bacterial colonies. *Journal of The Royal Society Interface*, 12(104):20141290–20141290, 2015.

[54] Hiroshi Fujikawa. Diversity of the growth patterns of bacillus subtilis colonies on agar plates. *FEMS microbiology ecology*, 13(3):159–167, 1994.

[55] T. A. Witten and L. M. Sander. Diffusion-limited aggregation. *Physical Review B*, 27(9):5686–5697, 1983.

[56] Hiroshi Fujikawa and Mitsugu Matsushita. Fractal Growth of Bascillus Subtilus on Agar Plates. *Journal of the Physical Society of Japan*, 58(11):3875–3878, 1989.

[57] Ferdi L Hellweger, Robert J Clegg, James R Clark, Caroline M Plugge, and Jan-Ulrich Kreft. Advancing microbial sciences by individual-based modelling. *Nature Publishing Group*, 14, 2016.

[58] Isaac Klapper and Jack Dockery. Finger formation in biofilm layers. *SIAM Journal on Applied Mathematics*, 62(3):853–869, 2002.

[59] Masayasu Mimura, Hideo Sakaguchi, and Mitsugu Matsushita. Reaction–diffusion modelling of bacterial colony patterns. *Physica A: Statistical Mechanics and its Applications*, 282(1):283–303, 2000.

[60] M. Matsushita, J. Wakita, H. Itoh, I. Ràfols, T. Matsuyama, H. Sakaguchi, and M. Mimura. Interface growth and pattern formation in bacterial colonies. *Physica A: Statistical Mechanics and its Applications*, 249(1):517–524, 1998.

[61] So Kitsunezaki. Interface Dynamics for Bacterial Colony Formation. *Journal of the Physical Society of Japan*, 66(5):1544–1550, 1997.

[62] Jan-Ulrich Kreft, Cristian Picioreanu, Mark C. M. van Loosdrecht, and Julian W. T. Wimpenny. Individual-based modelling of biofilms. *Microbiology*, 147(11):2897–2912, 2001.

[63] M. Matsushita, F. Hiramatsu, N. Kobayashi, T. Ozawa, Y. Yamazaki, and T. Matsuyama. Colony formation in bacteria: experiments and modeling. *Biofilms*, 1(4):305–317, 2004.

[64] Deborah Schwarcz, Herbert Levine, Eshel Ben-Jacob, and Gil Ariel. Uniform modeling of bacterial colony patterns with varying nutrient and substrate. *Physica D: Nonlinear Phenomena*, 318:91–99, 2016.

[65] JB Xavier, C Picioreanu, and MCM Van Loosdrecht. A modelling study of the activity and structure of biofilms in biological reactors. *Biofilms*, 1(4):377, 2004.

[66] Joao B Xavier, Cristian Picioreanu, and Mark CM Van Loosdrecht. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environmental microbiology*, 7(8):1085–1103, 2005.

[67] William PJ Smith, Yohan Davit, James M Osborne, Wook Kim, Kevin R Foster, and Joe M Pitt-Francis. Cell morphology drives spatial patterning in microbial communities. *Proceedings of the National Academy of Sciences*, 114(3):E280–E286, 2017.

[68] William D Hamilton. The genetical evolution of social behaviour. ii. *Journal of theoretical biology*, 7(1):17–52, 1964.

[69] Knut Drescher, Carey D Nadell, Howard A Stone, Ned S Wingreen, and Bonnie L Bassler. Solutions to the public goods dilemma in bacterial biofilms. *Current Biology*, 24(1):50–55, 2014.

[70] Robin A Sorg, Leo Lin, G Sander Van Doorn, Moritz Sorg, Joshua Olson, Victor Nizet, and Jan-Willem Veening. Collective resistance in microbial communities by intracellular antibiotic deactivation. *PLoS biology*, 14(12):e2000631, 2016.

[71] Matthew Simmons, Matthew C Bond, Knut Drescher, Vanni Bucci, and Carey D Nadell. Evolutionary dynamics of phage resistance in bacterial biofilms. *bioRxiv*, page 552265, 2019.

[72] Philip J Gerrish and Richard E Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102:127, 1998.

[73] James D Murray. *Mathematical biology: I. An introduction*, volume 17. Springer Science & Business Media, 2007.

[74] Laurent Excoffier, Matthieu Foll, and Rémy J Petit. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40:481–501, 2009.

[75] Sara Mitri, Ellen Clarke, and Kevin R Foster. Resource limitation drives spatial organization in microbial groups. *The ISME journal*, 10(6):1471–1482, 2016.

[76] Jordi Van Gestel, Franz J Weissing, Oscar P Kuipers, and Akos T Kovács. Density of founder cells affects spatial pattern formation and cooperation in bacillus subtilis biofilms. *The ISME journal*, 8(10):2069–2079, 2014.

[77] Venhar Celik Ozgen, Wentao Kong, Andrew E Blanchard, Feng Liu, and Ting Lu. Spatial interference scale as a determinant of microbial range expansion. *Science advances*, 4(11):eaau0695, 2018.

[78] Masatoshi Nei, Takeo Maruyama, and Ranajit Chakraborty. The bottleneck effect and genetic variability in populations. *Evolution*, pages 1–10, 1975.

[79] Sara Mitri, João B Xavier, and Kevin R Foster. Social evolution in multispecies biofilms. *Proceedings of the National Academy of Sciences*, 108(Supplement 2):10839–10846, 2011.

[80] Matti Gralka and Oskar Hallatschek. Environmental heterogeneity can tip the population genetics of range expansions. *Elife*, 8:e44359, 2019.

[81] N. Dalchau, P. K. Grant, J. R. Brown, A. Phillips, and J. Haseloff. Programming spatial patterns in bacterial colonies using engineered orthogonal signalling channels. In *IET/SynbiCITE Engineering Biology Conference*, pages 1–1, 2016.

[82] Fred D Farrell, Matti Gralka, Oskar Hallatschek, and Bartlomiej Waclaw. Mechanical interactions in bacterial colonies and the surfing probability of beneficial mutations. *Journal of The Royal Society Interface*, 14(131):20170073, 2017.

[83] Laurent A. Lardon, Brian V. Merkey, Sónia Martins, Andreas Dötsch, Cristian Picioreanu, Jan Ulrich Kreft, and Barth F. Smets. iDynoMiCS: Next-generation individual-based modelling of biofilms. *Environmental Microbiology*, 13(9):2416–2434, 2011.

[84] Cristian Picioreanu, M van Loosdrecht, and J Heijnen. *Multidimensional modeling of biofilm structure*. Delft University of Technology, Faculty of Applied Sciences, 1999.

[85] Cristian Picioreanu, Mark C M van Loosdrecht, and Joseph J Heijnen. Two-dimensional model of biofilm detachment caused by internal stress from liquid flow. *Biotechnology and Bioengineering*, 72(2):205–218, 2001.

[86] Achi Brandt. Multi-Level Adaptive Solutions to Boundary-Value Problems. *Mathematics of Computation*, 31(138):333–390, 1977.

[87] C. Picioreanu, J. B. Xavier, and M. C. M. van Loosdrecht. Advances in mathematical modeling of biofilm structure. *Biofilms*, 1(4):337–349, 2004.

[88] Joao B Xavier, Cristian Picioreanu, and Mark C M Van Loosdrecht. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environmental Microbiology*, 7(8):1085–1103, 2005.

[89] Thomas Bjarnsholt, Peter Østrup Jensen, Mark J. Fiandaca, Jette Pedersen, Christine Rønne Hansen, Claus Bøgelund Andersen, Tacjana Pressler, Michael Givskov, and Niels Høiby. Pseudonomas auriginosa biofilms in the respiratory tract of cystic fibrosis patients. *Pediatric Pulmonology*, 44(6):547–558, 2009.

[90] Haluk Beyenal, Suet Nee Chen, and Zbigniew Lewandowski. The double substrate growth kinetics of Pseudomonas aeruginosa. *Enzyme and Microbial Technology*, 32(1):92–98, 2003.

[91] Kasper N. Kragh, Jaime B. Hutchison, Gavin Melaugh, Chris Rodesney, Aled E.L. Roberts, Yasuhiko Irie, Peter Jensen, Stephen P. Diggle, Rosalind J. Allen, Vernita Gordon, and Thomas Bjarnsholt. Role of multicellular aggregates in biofilm formation. *mBio*, 7(2), 2016.

[92] Joseph A. Robinson, Michael G. Trulear, and William G. Characklis. Cellular reporoduction and extracellular polymer formation byPseudomonas aeruginosa in continuous culture. *Biotechnology and Bioengineering*, 26(12):1409–1417, 1984.

[93] R. Bakke, M. G. Trulear, J. A. Robinson, and W. G. Characklis. Activity of Pseudomonas aeruginosa in biofilms: Steady state. *Biotechnology and Bioengineering*, 26(12):1418–1424, 1984.

[94] Rubin Battino, Timothy R. Rettich, and Toshihiro Tominaga. The Solubility of Oxygen and Ozone in Liquids. *Journal of Physical and Chemical Reference Data*, 12(2):163–178, 1983.

[95] Philip S Stewart. Diffusion in biofilms. *Journal of bacteriology*, 185(5):1485–1491, 2003.

[96] Bruce E Rittmann and Jacques A Manem. Development and experimental evaluation of a steady-state, multispecies biofilm model. *Biotechnology and Bioengineering*, 39(9):914–922, 1992.

[97] Erik Alpkvist, Cristian Picioreanu, Mark C.M. van Loosdrecht, and Anders Heyden. Three-dimensional biofilm model with individual cells and continuum EPS matrix. *Biotechnology and Bioengineering*, 94(5):961–979, 2006.

[98] Cristian Picioreanu, Mark C. M. van Loosdrecht, and Joseph J. Heijnen. Mathematical modeling of biofilm structure with a hybrid differential-discrete cellular automaton approach. *Biotechnology and Bioengineering*, 58(1):101–116, 1998.

[99] Silvia N Santalla and Silvio C Ferreira. Eden model with nonlocal growth rules and kinetic roughening in biological systems. *Physical Review E*, 98(2):22405, 2018.

[100] A Brú and D Casero. Anomalous scaling of multivalued interfaces. *EPL (Europhysics Letters)*, 64(5):620, 2003.

[101] Belén Moglia, Ezequiel V Albano, and Nara Guisoni. Pinning-depinning transition in a stochastic growth model for the evolution of cell colony fronts in a disordered medium. *Physical Review E*, 94(5):52139, nov 2016.

[102] Knut Drescher, Jörn Dunkel, Carey D Nadell, Sven Van Teeffelen, Ivan Grnja, Ned S Wingreen, Howard A Stone, and Bonnie L Bassler. Architectural transitions in vibrio cholerae biofilms at single-cell resolution. *Proceedings of the National Academy of Sciences*, 113(14):E2066–E2072, 2016.

[103] David Head. personal communication, 2017.

[104] Chiara Giverso, Marco Verani, and Pasquale Ciarletta. Emerging morphologies in round bacterial colonies: comparing volumetric versus chemotactic expansion. *Biomechanics and Modeling in Mechanobiology*, 15(3):643–661, 2016.

[105] Pahala Gedara Jayathilake, Prashant Gupta, Bowen Li, Curtis Madsen, Oluwole Oyebamiji, Rebeca González-Cabaleiro, Steve Rushton, Ben Bridgens, David Swailes, Ben Allen, A. Stephen McGough, Paolo Zuliani, Irina Dana Ofiteru, Darren Wilkinson, Jinju Chen, and Tom Curtis. A mechanistic Individual-based Model of microbial communities. *PLOS ONE*, 12(8):e0181965, 2017.